LECTURES ON THE
SCIENTIFIC BASIS OF MEDICINE

1951–52

British Postgraduate Medical Federation
University of London

LECTURES ON THE SCIENTIFIC BASIS OF MEDICINE

Volume I

1951–52

UNIVERSITY OF LONDON
THE ATHLONE PRESS
1953

Published by
THE ATHLONE PRESS
at the Senate House, London, W.C.1
Distributed by Constable & Co. Ltd.
12 Orange Street, London, W.C.2

Printed in Great Britain by
WESTERN PRINTING SERVICES LTD.
BRISTOL

PREFACE

It is difficult for anyone engaged in the close study of a branch of medicine or following a career in one of the sciences on which the practice of medicine is based to find time to read exhaustively in fields of knowledge other than the one in which he is himself actively engaged. He welcomes, therefore, any help that enables him to realize the fundamental advances that are being made in other fields and their possible influence on his own work. The series of lectures on 'The Scientific Basis of Medicine' arranged each winter by the British Postgraduate Medical Federation is designed to supply such help, and the audiences that consistently attend these lectures are evidence of their success. The subjects are chosen because they illustrate the methods by which new and fundamental knowledge is being acquired, and the lecturers are chosen because they are themselves contributing to this new knowledge Some of the lectures contain fresh observations which are best published immediately in scientific journals so that they may be subjected to informed criticism, and some others consist mainly of matter that has been recently published elsewhere. The remainder bring up to date the progress that is being now made in a subject against a background of the history of its earlier advances, and explain the newer methods by which the progress has been possible. It is the lectures in this last group that are especially suitable for publication in a volume where they can be studied at leisure; they provide critical statements of the basic knowledge in their subjects from which fresh investigations can be planned or which can be applied to practical medicine after critical study in the wards and clinical laboratories. It is unfortunate that during the inevitable interval between the preparation of a lecture and its publication important advances may well occur, particularly in those subjects in which development

is most active. The dates of actual delivery of the lectures in this volume have therefore been given at the end of this Preface.

Of the thirty-nine lectures delivered during the winter of 1951–2, eighteen are included in this first annual volume, and it is hoped to produce a similar volume of suitable lectures in each succeeding year. Like the lectures themselves the volumes are intended especially for the younger research workers and teachers in the pre-clinical and clinical sciences and for junior clinicians who hope to find careers as consultants and teachers in the various special branches of medicine and surgery.

The British Postgraduate Medical Federation and the committee of its Central Academic Council which is directly responsible for arranging the series of lectures and selecting those included in this volume are indebted to the officers of the Athlone Press for assistance in preparing the material for publication.

FRANCIS FRASER
Director, British Postgraduate Medical Federation

3 November 1952

NOTE

The lectures printed in this volume were delivered on the following dates

I	9 October 1951	X	23 November 1951
II.	16 November 1951	XI	27 November 1951
III	21 February 1952	XII.	20 November 1951
IV.	17 January 1952	XIII.	11 December 1951
V.	4 March 1952	XIV	7 February 1952
VI	28 February 1952	XV	26 February 1952
VII	31 January 1952	XVI	12 February 1952
VIII.	7 November 1951	XVII	16 October 1951
IX	11 October 1951	XVIII.	20 March 1952

CONTENTS

I

The Scientific Approach to Medical Research

E. D. ADRIAN

MOST of us were at school when we decided to become doctors and no doubt we believed that the chief reason for our choice was a praiseworthy desire to look after our fellow men when they were sick At that age one is not ashamed of generous impulses and has not learnt to question their origin. Few of us now would be bold enough to assert that we can always decide rationally and from the best of motives; we may suspect that we took to medicine to please or spite our parents or because we wanted power over weaker people or for the more prosaic reason that circumstances conspired to push us into it. Yet the fact that there is a good reason for medical work is not to be set aside, and most of us can take comfort in the reflection that in this case at all events our unconscious urges made us choose very sensibly.

We have had or are having a long training in the theory and practice of medicine. There is a vast body of knowledge for us to learn and a great deal of successful treatment for us to give. We can expect to do some good to our fellows, if we learn this and spend all our time applying what we know. But for many of us a training in medicine is a frustrating experience because there is so much to learn that is speculative and so little that is quite certain. Even the student who believes all he is told and all he reads in the textbook cannot fail to realize that he will be told differently tomorrow, that theory and treatment change rapidly and seem never to reach finality. There are accepted principles

but an appendix or a fracture is not treated in precisely the same way in all our hospitals and the worldly-wise student still thinks it advisable to study the personal opinions of his examiners. This does not matter so much when our aim is only to pass an examination but it is disturbing to have no certainty when we come to treat the sick. Clinical practice may satisfy our social instincts when we can cure the patient, but it cannot avoid bringing a sense of failure when we do not know what to do; not so much of personal failure as of the failure of the whole race of doctors and of their teachings. All of us then want less speculation and more certainty. All of us are driven by the nature of our calling to try to discover more about sickness and health, to become investigators as well as practitioners.

Nowadays too our calling forces us to face new problems. Our main trouble must be that patients still die prematurely, but medicine really has advanced so much in the last hundred years that we have begun to see that our traditional aims are too narrow for what we can do. Physicists may be blamed because they have learnt to disintegrate the atom, but medicine may lead to far more trouble by saving lives than they are likely to do by destroying them. It is our science which has made it possible for the population of the world to increase more rapidly than its food supplies and we ought not to be blind to the consequences. And another thing that is forcing our hands is the realization that there is no hard and fast line between health and sickness when we are dealing with the whole of human activity. All of us have stresses and anxieties which can make us behave badly, and if there is a chance that medical knowledge can make people behave better than they do, we must not be too preoccupied with saving their lives to the exclusion of their happiness and their relations to one another.

So our calling was a good one to have chosen but we cannot escape all the problems it brings with it. We can try to solve them in various ways—by observation at the bedside, by statistical studies of large numbers, or by work in a laboratory. All these methods involve a considerable background of knowledge and a rational approach to the problem. Most of them involve the approach which is based on the background of present-day

natural science. This is the approach which can give the most satisfying answers and I suppose most of us would like to use it if we could.

Although it comes into the title of this lecture I do not much like the word scientific applied to medical research or to anything else. Often it means no more than rational as opposed to irrational and when we say that a theory is unscientific we simply mean that we think it bad. But there is one sense in which 'scientific' can be used with a more definite meaning, the meaning that our ideas are based on the general assumptions of the physical sciences and that our aim is to fit our observations into the picture of the material world which has been elaborated in the past three hundred years, the picture of electrons and atoms and molecules which the chemists and physicists have used so well for their own data. The scientific attitude in this sense is one in which all the diversity in the appearance of things is to be thought of as due to the mathematical gyrations of a few different kinds of elementary particle, and the scientific approach to medical research is that of the biochemist and biophysicist whose aim is to bring all the activities of living organisms into this scheme. It is of course a completely materialistic scheme, but its failings can be left to the philosophers. Whatever its ultimate truth there is no doubt that it is a convenient method of sorting out our observations. It is a handy way of looking at things and deciding where to look next.

Now there are special difficulties about applying the standard conceptions of natural science and the standard methods of advance when we are trying to understand human disease We are dealing with biological material, living bodies which are structures organized in such a way that they have properties unlike those of inanimate matter. We share that difficulty with biologists generally, but our biological material is the most difficult of all. It is not merely alive but it is as conscious and intelligent as we are. And since our material is man we cannot quite avoid approaching it with a slightly emotional bias Although patients may be constructed of the same stuff as lumps of coal and drops of water it is difficult to think quite dispassionately of their simple origins when they are in trouble. For much the same

reasons our social feelings and our social organization prevent us from undertaking the kind of experiment which would make one patient worse even though it might tell us how to make others better. Without experiments to show whether we are on the right track, with our emotions engaged and with patients who have minds as well as bodies it is no wonder that we are not always helped by considering the scientific picture of the natural world.

But the picture becomes more and more useful to contemplate as we go further from the bedside and nearer to the laboratory, as we think less about the patient as a human being and more about him as a collection of cells. It is most useful when instead of the cells we think of the enzymes and colloidal networks and interfaces, though by that time we are usually too far away from the patient to take more than an academic interest in him or his illness. When his feelings have to come into the picture it ceases to be that of mathematical physics.

Now we cannot all keep in the laboratory and I think it is worth insisting at the start that research need not be less rational or less successful when it is concerned with data which have not yet been fitted into what Isaac Newton called 'the frame of the system of the World'. At all events there are aspects of human activity which do not fit easily into this frame; they are aspects which we are bound to consider and we do manage to learn more about them although we cannot reduce them to mathematical treatment. It may be true to say that knowledge becomes more scientific in proportion to the amount of mathematics in it, but there is a great deal of knowledge about the natural world which is not mathematical but is still worth having.

However there is a great deal to learn about the physical and chemical properties of the body in health and disease and these properties can be studied by the all-powerful methods of natural science. The chief trend of medical research nowadays is to do so and to produce more and more exact and quantitative information. The advance of science generally has put new kinds of measuring instruments into our hands and every day the range of medical information increases. First there were measurements of the temperature and of the blood pressure, records of

the heartbeat, mechanical and then electrical; now there are records from the brain and as well as many more physical measurements we have all the resources of the biochemist to give us quantitative data about the patient's molecular constitution.

I must speak as an observer from outside and all I can do is to discuss some of the special features which are likely to appear in medical research even when it is confined to this quantitative and analytical plane, the features which make it so different from research in physics or chemistry or even in veterinary medicine. There is a whole profession anxious to take part in it; we cannot afford to be wasting our time. The suspicion that some people may be wasting their time is due mainly to the great variety of measurements that can now be made. With some of these it is not yet certain how far the measurements can vary in a healthy person and all that can be said is that they can now be made. Is there any ground for the belief that our investigators often measure or record something merely because they have found a way of doing it and not because they see clearly how it might yield further information? Of course there is. If they are not anxious to see what they can get they must be very unenterprising people and they deserve to miss an important discovery. But although they do not waste their time the consequences may be troublesome because so many people would like to help in medical research. The investigator or the team will naturally wish to explore the possibilities of a new quantitative test or a new kind of electrical record on patients of all kinds and before very long those who have to deal with the patients will think of the new test as one which ought to be made in any up-to-date hospital The application of the test will no longer be confined to the research team where it began and where its value can be assessed, but it will become diffused throughout the hospital world so that doctors who are in doubt about a patient will ask for it to be done to satisfy the natural desire to do everything possible. The pathologists and biochemists will find that their time is taken up with measurements of uncertain value in which they are not specially interested and the final result may well be that the work is turned over to specially trained technical experts who are the last people to give a dispassionate

judgment on the value of what they are doing. I know this is an exaggeration, but it takes a bold man to maintain that the data he can furnish cannot be of some value, if not to a particular patient, at least to future investigators of the disease. The point, surely, is that if such data are to be valuable there must be the right people to consider them.

We have to face the fact that in medicine research cannot often proceed by a direct frontal attack. The scientist who works in a laboratory can pick and choose his material so that he can experiment with only one variable, and his success depends on the skill with which he can fix all the other conditions. He tries to avoid collecting a mass of data which may be irrelevant and when he works on inanimate matter he can usually do so. And usually his results come quickly so that he can see whether he is on the right track. Accidental observations may help, but his work has a definite direction from the start. In medical research we cannot make our own conditions or decide where to look. Unless the disease or something like it can be produced in animals we can only observe, and the tempo of events is often so slow that we can get little guidance from observations on a single individual. All sorts of changes may intervene to obscure the issue and the slower the disorder the more of them there will be. Besides suffering from his malady the patient may rest or change his occupation or his diet and he will certainly grow older. With so much that may or may not be relevant we may be forced to pile up all possible observations and measurements in the hope of something turning up

This is not a fruitful or an elegant method of research and it is not the chosen method of natural science, but it is a method which had at one time a very powerful advocate, no less than the lord chancellor of England, Francis Bacon. Bacon advocated fact-finding on a grand scale conducted by mighty institutes filled with observers taking notes. There was to be some method in their labours and they were encouraged to make experiments, but one experiment was as good as another. If only enough data were collected and then sorted out dispassionately on a rigid system Bacon held that the plan of nature could not fail to emerge to the inestimable benefit of mankind.

Bacon was a philosopher who knew a great deal about human nature. He saw well enough how men can deceive themselves when they leave fact for theory and this may have pushed him to the other extreme in not allowing the search for facts to be directed. As Broad says, he never clearly distinguished between approaching facts with a prejudice and approaching them with a working hypothesis. But at the time he wrote the prejudices were still in the ascendant. His plan for collecting the facts of nature into an immense history and then letting them speak for themselves was the first attempt to lay down the principles on which new discoveries were to be made and the method seemed the more certain in that it made no great demands on the intelligence or imagination of the observers. They were to be plain men subsidized by the great. They had only to produce enough observations and the general laws uniting them would become clear when they were put through the inductive machinery ✓which Bacon was devising. It promised a straight road to the conquest of nature.

Unfortunately Bacon was a busy man and never managed to give more than a preliminary sketch of the rules by which the facts were to be sorted. We can approve his eminent good sense in practical affairs—for instance his demands that university lecturers should have higher salaries and money for apparatus—but the examples he gives of his inductive method in operation are not very encouraging. And although his writings had an influence on European thought which lasted a century or more there was no attempt to put the great plan into action. Discoveries came rapidly but they were made by men who had little use for such comprehensive and automatic research. Boyle admired Bacon's writings but he and Harvey worked quite differently, with fewer observations and more imagination. Presumably the Baconian method was never tried because it was never thoroughly explained and because the inquiring minds of those days were not willing to follow a road which needed dogged persistence instead of curiosity and intelligence—even though they had Bacon's assurance that the road would lead to great achievement.

Now there is still a good deal of medical research in which

something rather like the Baconian method has to be used. In default of experiments we must make observations and with human material deliberate experiments are often impossible. You cannot make housing conditions worse to see if there is a corresponding increase in tuberculosis, but if someone has recorded housing conditions and someone else has recorded the incidence of tuberculosis, the machines that sort out the punched cards should be quite capable of putting two and two together. And so a collection of all possible data about the body may be valuable when they are sorted out and if they are expressed in a quantitative way there will be far more chance of sorting them effectively.

The chief trouble with this kind of approach is that it must be very dull work, at all events in the initial stages. Bacon thought of a large body of investigators but to keep a large team collecting facts without meaning would have needed a different leader with a much less critical and cautious temperament. And in medical research there is again the difficulty that the whole training of a doctor makes it hard to contemplate such an impersonal approach

But it would be a great mistake if this line of research were to cease. When the present era of medicine began its rewards were clear enough. Koplik could discover his spots and Addison the lesions in the suprarenal gland and now although the ground has been worked over by such acute observers we have all the fresh data given by exact measurements. Unless they are to be completely buried the labour spent in making them will not be thrown away. It is easy to encourage rather tedious routine work when one does not have to do it oneself, but the harassed house physicians and registrars who have to write the elaborate case records, and the pathologists and biochemists who add to them, must be made to realize that they are really contributing to the advance of knowledge.

I can remember when I was a resident at Queen Square how reluctant I was to believe that it was really important to make such long notes on every case. I suspected that they were kept mainly because it was a part of the accepted tradition of the hospital. I had the conscious superiority of a young man fresh

from experimental work and it took all Dr. Walshe's persuasiveness to convince me that I might be mistaken. A major factor in my conversion was the paper which Kinnear Wilson had pub- [illegible]

the liver and degeneration of the lenticular nucleus of the brain. The curious assortment of symptoms and pathological changes caught Wilson's interest. He searched the published records. Gowers had described two patients with the same clinical history, Ormerod one, and Homen of Helsingfors three. The notes made by Gowers and Ormerod were at Queen Square and were detailed enough to make it clear that Wilson's patient had the same disorder. He searched for further cases and found three more in mental hospitals, and the story ends with a hurried journey to Switzerland where he came just in time to attend the post mortem on another patient whose body had been kept high up in the mountains to preserve the brain against his arrival.

Wilson's disease was established as a clinical and pathological entity by this paper, though nowadays the relations he traced have been detected in a wider field. But the progress of knowledge in this instance shows how the method of observation can sometimes be a much more fruitful way of advance than the method of experiment. Wilson made the reasonable decision that he should try to extend his ideas by producing experimental lesions in the lenticular nucleus of animals to see how the muscular system would be affected. He carried out a long series of experiments on monkeys. They were well planned in the light of existing knowledge about the basal ganglia, but the results were entirely negative and threw no light at all on the mechanism of the disease. No one who knew Wilson's great ability can read this second paper about his experimental work without regret that it gave him no opportunity to go on· and no one can read his first paper without enjoyment at the way in which he seized all the opportunities provided by his own observations and those made by others which were accessible to him.

One could multiply stories like that of Wilson's disease to

show the value of full records and this particular story does not involve any of the exact measurements which would be recorded nowadays. I have dragged it in to illustrate the rather obvious point that in addition to the records there must be someone interested enough to examine them. Sorting machines may help but only when there is someone with the urge to search out fresh relations and the ability to see them.

We must be careful therefore that too much insistence on the scientific attack does not discourage our fact collectors by making them feel that they have a profitless task. It must somehow be made clear that the records and measurements they are called on to make are not entirely wasted when they do not happen to assist the diagnosis of a particular case. The fact that our material is human sickness must not be allowed to make us unduly anxious for quick results. It may well be irksome to follow an inquiry which can scarcely benefit a particular patient though in the long run it may benefit many, and so we must make sure that the long-term benefits are not lost through lack of interest.

I said that our material is human sickness and I might have added human suffering. But we have been thinking of the scientific approach to medical research with its insistence on the material system of the body. The framework of natural science was not designed to include human thoughts and feelings and [illegible]

and conscious. And the framework of mathematical physics is not inexpansible. Its boundaries are the constant sport of the cosmologists and we can be quite sure that when it becomes profitable to do so it can be made to include not merely the human mind but even the human mind engaged in such awkward activities as telepathy and precognition. At the present stage of natural science however, it is not yet profitable to approach *intelligent behaviour in the narrowly mechanistic sense* unless we are doing so intentionally as brain physiologists.

It is scarcely profitable because the special properties of the brain are there to defeat us. We are endowed with this remark-

able organ for synthesizing past and present experience, for making the body react to external events in a way which depends on a long train of past circumstance and seems to involve a stage of the peculiar kind of activity which we call consciousness. We cannot yet come anywhere near to a mathematical treatment of what goes on at this level and until we are able to do so we must be content with a much less formal approach when we have to consider what the patient may feel, think and do about his illness.

The special problems of mental disorder are scarcely relevant. It is reasonable to try to solve them by examining all the material properties of the body in the hope of finding some abnormality which has affected the brain. The appeal to experiment is no doubt more difficult than it is with bodily disease, for experiments on animals can be little guide when we are dealing with disorders which are peculiar to man because they are disorders of the human brain which is our peculiar possession. But until [illegible]

What I am concerned with is not the effect of the body on the mind, or if you prefer it on conscious activity or on the higher levels of the brain, but rather the possible effects of the mind or brain on the body.

Considerations of this kind scarcely arise in the more acute forms of illness. When the body is thrown out of gear by a sudden accident or infection the brain may be affected too but in any case it can do little to modify the course of events. The patient will decide to go to bed and send for a doctor and no doubt these decisions will have an effect but there will be little opportunity to build up an elaborate system of ideas which might have a more direct action on the body. But in chronic illness there is time to think, time that is to say for the disturbance in the body to be mirrored in the brain, submitted to the analysis based on past experience, given the appropriate emotional colouring and built up into a highly organized mental picture.

Our training in physiology makes us too much inclined to think of the signals from the sense organs as producing an

show the value of full records and this particular story does not involve any of the exact measurements which would be recorded nowadays. I have dragged it in to illustrate the rather obvious point that in addition to the records there must be someone interested enough to examine them. Sorting machines may help but only when there is someone with the urge to search out fresh relations and the ability to see them.

We must be careful therefore that too much insistence on the scientific attack does not discourage our fact collectors by making them feel that they have a profitless task. It must somehow be made clear that the records and measurements they are called on to make are not entirely wasted when they do not happen to assist the diagnosis of a particular case. The fact that our material is human sickness must not be allowed to make us unduly anxious for quick results. It may well be irksome to follow an inquiry which can scarcely benefit a particular patient though in the long run it may benefit many, and so we must make sure that the long-term benefits are not lost through lack of interest.

I said that our material is human sickness and I might have added human suffering. But we have been thinking of the scientific approach to medical research with its insistence on the material system of the body. The framework of natural science was not designed to include human thoughts and feelings and we must be prepared to approach them in a less rigid spirit. We need not suppose that there should be any hard and fast separation of material objects into non-living and living, unconscious and conscious. And the framework of mathematical physics is not inexpansible. Its boundaries are the constant sport of the cosmologists and we can be quite sure that when it becomes profitable to do so it can be made to include not merely the human mind but even the human mind engaged in such awkward activities as telepathy and precognition. At the present stage of natural science however, it is not yet profitable to approach intelligent behaviour in the narrowly mechanistic sense unless we are doing so intentionally as brain physiologists.

It is scarcely profitable because the special properties of the brain are there to defeat us We are endowed with this remark-

able organ for synthesizing past and present experience, for making the body react to external events in a way which depends on a long train of past circumstance and seems to involve a stage of the peculiar kind of activity which we call consciousness. We cannot yet come anywhere near to a mathematical treatment of what goes on at this level and until we are able to do so we must be content with a much less formal approach when we have to consider what the patient may feel, think and do about his illness.

The special problems of mental disorder are scarcely relevant. It is reasonable to try to solve them by examining all the material properties of the body in the hope of finding some abnormality which has affected the brain. The appeal to experiment is no doubt more difficult than it is with bodily disease, for experiments on animals can be little guide when we are dealing with disorders which are peculiar to man because they are disorders of the human brain which is our peculiar possession. But until we have searched exhaustively for material changes and found nothing there is no reason why the scientific approach to mental disorder should not be as valuable as it is in bodily disorder. What I am concerned with is not the effect of the body on the mind, or if you prefer it on conscious activity or on the higher levels of the brain, but rather the possible effects of the mind or brain on the body.

Considerations of this kind scarcely arise in the more acute forms of illness. When the body is thrown out of gear by a sudden accident or infection the brain may be affected too but in any case it can do little to modify the course of events. The patient will decide to go to bed and send for a doctor and no doubt these decisions will have an effect but there will be little opportunity to build up an elaborate system of ideas which might have a more direct action on the body. But in chronic illness there is time to think, time that is to say for the disturbance in the body to be mirrored in the brain, submitted to the analysis based on past experience, given the appropriate emotional colouring and built up into a highly organized mental picture.

Our training in physiology makes us too much inclined to think of the signals from the sense organs as producing an

exactly corresponding signal in consciousness. We are inclined to neglect the profound additions and transformations which the brain can impose on a straightforward sensory message. No doubt the more intense and urgent the signal the less the brain can do to it. A violent pain brings the protective reactions of the lower levels of the nervous system into play, but will not be transmuted by the brain, for severe pains are not guides to intelligent action and play little part in the constructions of the cerebral cortex. But the familiar discomforts of a long illness can arrive in consciousness linked with a whole system of memories and anxieties. When a sufferer from dyspepsia tells us that when he eats pork he is painfully aware of each lump as it arrives in his stomach we may think that he is dramatizing his sensations but that is what he feels.

The brain cannot avoid this kind of elaboration for that is its main function, and most of the elaboration goes on at unconscious levels. I have made no attempt to keep clear of all the errors which can creep in whenever we speak of the brain and the mind in the same breath. They are too obvious to need pointing out and I can only wish that I had time to avoid them by following the excellent lead given by J. Z. Young in his Reith lectures. But to come to conscious levels we can see what the brain, or if you prefer it the unconscious mind is always doing by considering how we remember past events. Professor Bartlett has written a fascinating study of the transformations and elaborations which take place when anything is seen or heard and subsequently recalled to memory. It is a fascinating book partly from the way in which it avoids the jargon of the schools and partly because it puts the physiologists in their place—which is a long way from the finished products of the mind. He showed that as a rule we remember a little about the original material that was presented to us and then construct a whole picture according to our particular interests or social background. The finished picture has elements of the original material but it is a construction, a work of art rather than the development of a photograph. To the subject it often does not seem like this. He thinks he is remembering when he is really inventing and what he invents depends on who he is, how he was brought up, what he does for

his living and what sort of company he keeps. It is of course a familiar and chastening experience for a lecturer to see the distorting process at work when his students sit for an examination and he reads their version of his own teaching so curiously changed. Bartlett points out how the constructiveness which selects and colours what we remember is a development associated with consciousness, but has its roots in the process by which elementary sensations of all kinds are organized by the central nervous system. Henry Head pointed out how they are organized to guide any movement in relation to past movement and to the whole posture of the body, and we know something now of the very elaborate signalling apparatus in the muscles which help to give us our picture of the body and operate as a feedback system for adjusting the fresh movement.

What I have stressed is nothing more than the obvious point that each of us may feel a disorder of his body in an individual way because what we feel will be inevitably coloured by the way we have learnt to organize our experiences. In chronic illness therefore the discomforts and disabilities which the patient suffers will depend on what his brain does to the material which reaches it, will depend that is to say on his whole character and outlook and past experiences. This may have little effect on the course of the illness, but we are coming to realize that it may affect it considerably in ways of which we are still very little aware. Certainly we are rapidly learning from studies of the physiology of the brain in animals that its effects on the body may not be confined to the control of voluntary movement. It may well have profound effects on the autonomic system and on the endocrine glands. It is not unreasonable therefore to suppose that the whole course of a chronic disorder may be decided by the particular appeal that it makes to the brain or the mind of a particular patient.

This is not the familiar, though doubtless admirable, plea that one should consider the whole patient, the plea for the psychosomatic approach to his illness. It is rather a plea for directing some medical research to the possible interactions between the brain and the body. Some of it can be strictly scientific, experimental physiology aimed at further studies of the way in which

the body might be influenced by the brain, whether the brain can make a joint swell or a duodenum ulcerate, and some of it must be on the less secure basis which we have to use when we deal with the finished products of the brain, the thoughts and desires of the sick person.

I have strayed too far from my main topic, which was to consider some of the special features that are imposed on medical research by [illegible] too much ab[illegible] advances tha[illegible] made by the application of scientific methods. But I can at least illustrate the pace of this advance by recalling the time when I was a medical student at Cambridge. It is true that it was forty years ago, but modern methods were well under way. Lewis had begun his analysis of cardiac irregularities with the string galvanometer, and Hopkins was in Cambridge making biochemistry a science. We were naturally impressed with these developments—an electrocardiogram was such a change from the old routine way of examining a patient by taking his temperature and listening to his heart sounds. Now our Regius Professor of Physic was Clifford Allbutt. He was over seventy-five but still full of energy. He could hold an audience with a polished lecture, and though rather deaf, he presided most effectively at the faculty board of medicine. Medical students take their professors for granted and it was enough for us that Allbutt was an impressive figure, the editor of the standard textbook of medicine and of a guide to the preparation of an M.B. thesis. We knew he had climbed the Alps when he was younger and we were proud of him as one of the early generation of mountaineers But we did not know and I think we should have found it hard to realize that it was Allbutt who started the practice of taking the temperature. He did it in 1867 by getting a firm at Leeds to make the short clinical thermometer which we use now Before that a few inquiring physicians had measured temperatures with thermometers a foot long but most of them judged temperature by the feel of the patient's skin. It was our own professor who had shown the possibility and the value of precise temperature measurement and he had also shown the value of examining the eye with the ophthalmoscope.

I can claim therefore to have sat at the feet of the man who made us all aware of the scientific approach to medical research. He never referred to what he had done—at least not to us—and he might perhaps have been shocked if it had been suggested that he was partly responsible for the tendency to prefer instrumental records to direct observation. But I like to think that it was not so long ago, and with all its special difficulties the practice of medicine has certainly advanced an immeasurable distance since Clifford Allbutt set the fashion for the exact measurements which are the basis for the scientific approach to medical research.

the body might be influenced by the brain, whether the brain can make a joint swell or a duodenum ulcerate, and some of it must be on the less secure basis which we have to use when we deal with the finished products of the brain, the thoughts and desires of the sick person.

I have strayed too far from my main topic, which was to consider some of the special features that are imposed on medical research by the special nature of its material. And I have said too much about the difficulties and too little about the immense advances that have been made by the application of scientific methods. But I can at least illustrate the pace of this advance by recalling the time when I was a medical student at Cambridge. It is true that it was forty years ago, but modern methods were well under way. Lewis had begun his analysis of cardiac irregularities with the string galvanometer, and Hopkins was in Cambridge making biochemistry a science. We were naturally impressed with these developments—an electrocardiogram was such a change from the old routine way of examining a patient by taking his temperature and listening to his heart sounds. Now our Regius Professor of Physic was Clifford Allbutt. He was over seventy-five but still full of energy. He could hold an audience with a polished lecture, and though rather deaf, he presided most effectively at the faculty board of medicine. Medical students take their professors for granted and it was enough for us that Allbutt was an impressive figure, the editor of the standard textbook of medicine and of a guide to the preparation of an M.B. thesis. We knew he had climbed the Alps when he was younger and we were proud of him as one of the early generation of mountaineers. But we did not know and I think we should have found it hard to realize that it was Allbutt who started the practice of taking the temperature. He did it in 1867 by getting a firm at Leeds to make the short clinical thermometer which we use now. Before that a few inquiring physicians had measured temperatures with thermometers a foot long but most of them judged temperature by the feel of the patient's skin. It was our own professor who had shown the possibility and the value of precise temperature measurement and he had also shown the value of examining the eye with the ophthalmoscope.

onwards. There followed during the next two to three decades evidence for the existence of thyroxine, insulin and the hormones of the anterior pituitary gland. All these hormones were obtained by methods which depended on the solubility of the active material in watery solution and it was perhaps natural to assume that a hormone, which must be liberated into the internal environment—an aqueous phase—would be most effectively extracted from its gland of origin by means of an aqueous medium. But we now realize that this assumption is not always sound and that adrenal steroids are indeed most satisfactorily prepared by means of lipid-soluble solvents.

Twenty years ago it was generally assumed, despite some evidence to the contrary derived from studies on the anterior pituitary gland, that for one endocrine gland there was one hormone. Thyroxine (or perhaps thyroglobulin) was the hormone of the thyroid gland while insulin was the hormone of the pancreatic islets and adrenaline that of the adrenal medulla. But from 1935 onwards a large number of crystalline substances were isolated from extracts of the adrenal glands, of which seven were active in maintaining the life of adrenalectomized animals (Fig. 1). Both androgens and oestrogens were also obtained. These active substances and many others besides were isolated and characterized chiefly by Reichstein and his colleagues in Basle in Switzerland, and by Kendall and his co-workers in the Mayo Clinic, Rochester, Minnesota. The demonstration by Reichstein in 1937 that adrenal substances were close relatives of the sterols, bile acids and sex hormones and were therefore 'steroids', was as aesthetically satisfying as it was practically valuable in pointing the way to artificial production of adrenal steroids from commonly occurring materials such as bile acids. In fact, in 1937 Steiger and Reichstein first prepared 11-deoxycorticosterone artificially from bile acids before the substance had been identified in an extract of adrenal tissue. For many years now this substance—or its acetate, often known as doca—has been a commercial product of therapeutic value in the treatment of Addison's disease. Whether or not it is secreted by the adrenal cortex under normal conditions is still open to doubt.

In addition to the seven characterized crystalline active adre-

II

Adrenal Hormones and ACTH

F. G. YOUNG

THE impact on medicine of cortisone and ACTH, dating from the observation of Hench and his colleagues in 1948 that these substances can alleviate in a striking fashion symptoms of rheumatoid arthritis, has been tremendous, and the clinical conditions in which some measure of success has followed the administration of one or other of these drugs are legion. But in a lecture devoted to the scientific basis of medicine we shall have little to say about therapy as a practical measure. The amount of published experimental work concerning the influence of adrenal hormones and ACTH on metabolic processes is very great, and in a one-hour lecture only a fraction of the evidence can be considered. In order to throw into perspective present views concerning this problem I shall first say a few words about the historical background.

Addison's disease was first described in 1849 and the fatal results of experimental adrenalectomy in animals were first investigated in 1856. But it was not until 1929 that a potent extract of the adrenal glands, capable of maintaining adrenalectomized cats indefinitely in good health, was prepared by Swingle and Pfiffner. Some of the reasons for eighty years of failure to prepare an effective extract are instructive (cf Young, 1951). For instance the recognition of adrenaline (and later noradrenaline) as secretory products of the adrenal medulla, which followed from the investigations of Oliver and Schafer in 1894, gave a great stimulus to the investigation of ductless glands, and the idea of chemical messengers was put on a firm general basis by the investigations and writings of Bayliss and Starling from 1902

probably 1–20 mg/kg body wt./24 hrs. If changes in the amount of the metabolic products of the adrenal steroids excreted in the urine can be relied upon as an index of the change in the rate of secretion of the steroids by the gland, then the rate of secretion of steroids by the gland may be increased many-fold, perhaps ten times or more, under conditions of stress or trauma.

It is of particular interest that although cortisone is so valu-

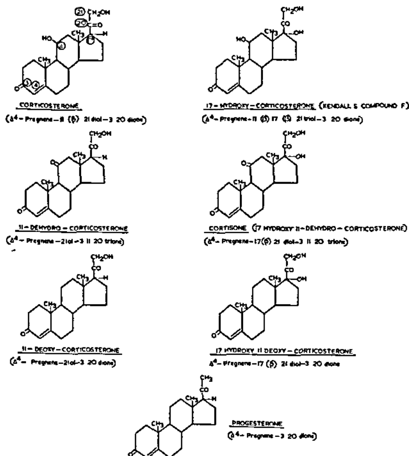


Fig. 1 Structures of the adrenal steroids active in prolonging the life of adrenalectomized animals

nal steroids there was also prepared from extracts of the adrenal gland an uncrystallizable syrup, named the amorphous fraction, which was particularly active in maintaining adrenalectomized animals in good condition. The properties of this fraction together with its chemical analysis suggest that it contains substances more highly oxygenated than the characterized steroids and these are of course more soluble in aqueous media.

NATURE OF THE SUBSTANCES SECRETED BY THE ADRENAL CORTEX

With the isolation of seven active materials and with the possibility that still more active substances might be obtained from the amorphous fraction there was naturally great discussion as to which, if any, was *the* hormone of the adrenal cortex. When corticosterone was isolated by Reichstein it was confidently named as such in the belief that it was the hormone. There is now little doubt that more than one substance is liberated into the blood by the cortex of the adrenal gland and that there is therefore more than one hormone.

The identification, accompanied in some cases by the isolation, of 17-hydroxycorticosterone together with corticosterone, in urine from Cushing's syndrome (Mason and Sprague, 1948), in urine from patients receiving treatment with ACTH (Mason, 1950; Sprague, Mason and Power, 1951), and in adrenal-vein blood (Reich, Nelson and Zaffaroni, 1950; Nelson, Samuels, Willardson and Tyler, 1951; Bush, 1951) together with the fact that 17-hydroxycorticosterone and corticosterone are produced in relatively large amounts by the isolated perfused adrenal gland (Hechter, Zaffaroni, Jacobsen, Levy, Jeanloz, Schenker and Pincus, 1951), allows us to conclude that these two steroids are the chief secretory products of the adrenal cortex. Bush (1951) has found that the proportion of the two steroids secreted under conditions of stress varies with the species of animal, but is relatively constant for each species. He suggests further that the adrenals of the dog may secrete two substances present in the 'amorphous fraction' under conditions which he describes as a 'steady state'.

The total amount of steroid secreted by the adrenal glands is

probably 1–20 mg/kg body wt./24 hrs. If changes in the amount of the metabolic products of the adrenal steroids excreted in the urine can be relied upon as an index of the change in the rate of secretion of the steroids by the gland, then the rate of secretion of steroids by the gland may be increased many-fold, perhaps ten times or more, under conditions of stress or trauma.

It is of particular interest that although cortisone is so valu-

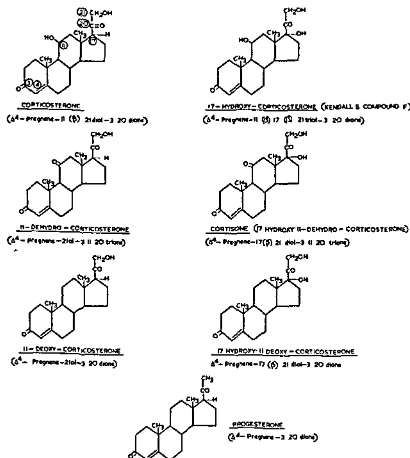


FIG. 1 *Structures of the adrenal steroids active in prolonging the life of adrenalectomized animals.*

able therapeutically this substance does not appear among the putative major secreted steroids of the adrenal cortex. Since, however, cortisone has been made available by commercial production from bile acids on a large scale, particularly by Merck and Co. in the United States, very much more is known about the biological activity of this substance than about that of either 17-hydroxycorticosterone or corticosterone, and much of what I shall have to say later will apply to cortisone rather than to what we now believe are the two major secretory products of the gland. The term 'adrenal steroid' will be used to describe steroids possessing an oxygen atom or a hydroxyl group at position 11-, chiefly 17-hydroxycorticosterone, corticosterone and cortisone. The term will not include 11-deoxycorticosterone unless this substance is specifically mentioned.

The presence of the 17-hydroxyl group appears to be essential for therapeutic value in rheumatoid arthritis and in rheumatic fever. Nevertheless the effectiveness of salicylate in the treatment of rheumatic fever, and indeed many similarities between the pharmacological effects of salicylate and cortisone, have led some to speculate as to whether salicylic acid, that is orthohydroxybenzoic acid, may mimic the activity of cortisone in the way that diethylstilboestrol mimics the naturally occurring oestrogens. Recently the remarkable effectiveness of 2:6-dihydroxybenzoic acid (γ-resorcylic acid) in rheumatic fever has been described by Reid, Watson, Cochran and Sproull (1951) in this country, and they have put forward the interesting suggestion that the mimicry of cortisone by salicylic acid and γ-resorcylic acid may depend on the ability of these substances to form chelate rings. Meta- and parahydroxybenzoic acids are ineffective in rheumatic fever and are unable to form chelate rings.

THE CONTROL OF THE SECRETORY ACTIVITY OF THE ADRENAL CORTEX

The importance of the influence of the anterior pituitary gland on the adrenal cortex was first revealed by the patient and elegant experiments of Philip Smith (1927, 1930). Smith showed that in the hypophysectomized rat the adrenal cortex undergoes substantial atrophy, but that this atrophy could be prevented or

cured by the subcutaneous implantation of rat anterior pituitary tissue. Since that time much research has been directed towards elucidating the nature of the adrenocorticotropic hormone (ACTH), liberated by the anterior pituitary gland, which is responsible for the maintenance of the adrenal cortex. The discovery by Long, Sayers and their colleagues that anterior pituitary extracts which could maintain the weight of the adrenal glands in hypophysectomized rats would also bring about a rapid fall in the ascorbic-acid content of the adrenal glands of the hypophysectomized rat, led to the development of a sensitive method for the assay of ACTH by Sayers, Sayers and Woodbury (1948). Another method that has been employed as a test for a rise in the amount of circulating adrenal steroids, particularly after the liberation of ACTH from the anterior pituitary [illegible]

assessing changes in the amount of circulating adrenal steroid have been of great value both experimentally and in clinical treatment.

ACTH

In 1943 Li, Evans and Simpson (1943) and independently Sayers, White and Long (1943) isolated, from sheep pituitaries and from pig pituitaries respectively, a protein which was believed to be pure and free from other anterior pituitary hormones and which I shall call protein-hormone ACTH. The two preparations of the protein-hormone exhibited similar physiological activities in the repair or maintenance of the adrenal glands in hypophysectomized rats, and also similar chemical and physico-chemical properties, including a molecular weight of 20,000. The two preparations were also equally effective in reducing the ascorbic-acid content of the adrenal cortex of the hypophysectomized rat.

In our laboratory we have recently succeeded, by means of ion-exchange resins, in showing that it is possible to remove from the protein-hormone a very small amount of basic material which contains nearly all the ascorbic-acid reducing activity of the preparation (Dixon, Moore, Stack-Dunne and Young,

1951). The protein left behind is virtually inactive (Fig. 2). Furthermore, we are able to separate to some extent on the ion-exchange column fractions which are active in the adrenal ascorbic-acid reducing test and fractions which are effective in

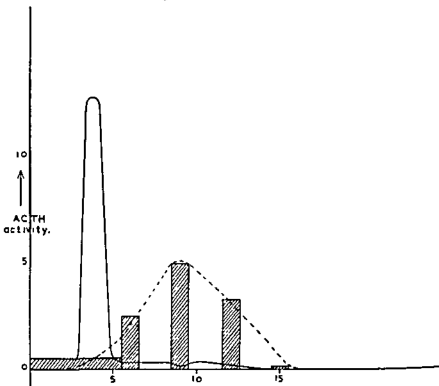


FIG. 2. A chromatogram of ACTH protein-hormone on Amberlite IRC-50 with sodium phosphate buffer of pH 6 8 The hatched blocks indicate activity measured by the ascorbic-acid reducing method of Sayers, Sayers and Woodbury (1948) The continuous line indicates the colour developed with ninhydrin under standard conditions. (Cf Dixon, Moore, Stack-Dunne and Young, 1951.)

the adrenal repair test in hypophysectomized rats (Dixon, Stack-Dunne, Young and Cater, 1951). We believe it to be very likely that ACTH is separated into at least two fractions without the need for a hydrolytic step, one fraction primarily concerned with the reduction of ascorbic acid in the adrenal gland, a function which may well be associated with the conversion of

the contained cholesterol to the adrenal steroids and the liberation of those that are secreted, and the other concerned with the maintenance of the adrenal cortex after hypophysectomy, a process which may concern the mechanism by which cholesterol is stored in the adrenal gland to await its subsequent conversion to adrenal steroids (Stack-Dunne and Young, 1951).

Li and his colleagues have described the preparation of peptides, obtained by the hydrolytic fission of the protein-hormone, which are active in various ACTH tests (cf. Reinhardt and Li, 1951). *Our own results appear to demonstrate that hydrolysis is* not needed to liberate the preformed basic material which is highly active in the ascorbic-acid reducing test and which is present in small proportion in the protein-hormone, and although this material may well be a peptide a rigid proof of this is still lacking.

An animal does not die as the result of hypophysectomy and in the hypophysectomized animal the adrenal cortex is not without function (Pickford and Vogt, 1951). There is evidence that the *zona glomerulosa* of the cortex is still active after removal of the pituitary gland, and capable of liberating sufficient steroid *to maintain the life of the hypophysectomized animal.*

Vogt (1951) has found that when the amount of potassium in the plasma of blood perfusing an isolated dog adrenal gland is raised, the output of steroids by the gland greatly increases. This occurs in the absence of the pituitary gland and shows that the adrenal cortex is capable of regulating its secretion to a limited extent independently of anterior pituitary ACTH.

Although anterior pituitary ACTH is not essential for minimal functioning of the adrenal cortex, for example, in the hypophysectomized animal, the activity of the cortex necessary for normal life in the intact animal, and particularly the rise in adrenal cortical activity which occurs under conditions of stress, are certainly dependent on the availability of ACTH from the anterior pituitary gland. The secretion of ACTH by the pituitary is itself regulated by three main mechanisms: (1) by the hypothalamus, (2) by the amount of adrenocortical hormones in the general circulation, and (3) by the amount of adrenaline and like substances in the blood.

Hypothalamic control of the secretion of ACTH. Groot and Harris (1950), using a method of stimulation of the hypothalamus of the unanaesthetized rabbit by means of electric currents induced by remote control, showed that stimulation of the posterior region of the *tuber cinereum* or of the mammillary body resulted in a fall in blood lymphocytes which was similar to that which followed the intravenous injection of a small dose of ACTH. Stimulation of other regions in the hypothalamus and pituitary gland was ineffective. Similar results have been obtained in the unanaesthetized dog by Hume and Wittenstein (1950). Conversely Harris found that the lymphopenia which normally follows stress was abolished if electrolytic lesions were placed in the hypothalamus, in the posterior part of the *tuber cinereum* or mammillary body, or in the zone of the anterior pituitary lobe close to the *tuber cinereum*. It was concluded that the secretion of ACTH by the anterior pituitary gland which is elicited by stress is regulated by the hypothalamus, and Harris believes that the stimulus is carried from the *tuber cinereum* to the anterior pituitary lobe by a humoral transmitter (or transmitters) which travels through hypophysial portal vessels (Harris, 1951). Not all workers accept the idea of humoral transmitters in this connexion and not all are agreed as to the significance of the hypophysial portal vessels.

The influence of blood adrenocortical hormone level on the secretion of ACTH Sayers (1950) has recently reviewed the evidence concerning the influence of adrenal steroids on the secretion of ACTH by the pituitary gland. Administration of crude adrenal extract or adrenal steroids to normal animals induces adrenal atrophy and will prevent the adrenal hypertrophy which normally follows stress. Since such effects are not seen in hypophysectomized animals it may be assumed that they result from a depression of secretion of ACTH by the anterior pituitary gland. Sayers believes that conditions such as stress increase the utilization of adrenal steroids by the peripheral tissues so that the blood level of these hormones falls, this fall leads the anterior pituitary gland to liberate greater amounts of ACTH. Conversely, an increase of blood adrenocortical hormone level brings about a depression of ACTH secretion. Sprague, Mason and Power (1951) have observed signs of adrenal insufficiency including asthenia

to follow cessation of cortisone therapy in human beings, and they consider these observations to support the view that cortisone depresses the secretion of pituitary ACTH.

The automatic regulation of ACTH secretion by the level of adrenal steroids in the blood seems clearly proved, though it is but one factor in a more complex mechanism for the maintenance of adequate amounts of adrenal steroids in the internal environment.

The influence of adrenaline and related substances on the secretion of ACTH Vogt (1944) observed that the intravenous infusion of adrenaline could cause an immediate and large increase in the rate of secretion of steroids by the adrenal cortex, but Long and Fry (1945) found that the influence of adrenaline on the adrenal cortex is very much less in hypophysectomized animals than it is in normal animals, and it appeared likely that adrenaline acts by stimulation of the secretion of ACTH by the anterior pituitary gland. More recently Vogt (1951) has shown that adrenaline does not increase the output of cortical-hormones by the isolated perfused dog adrenal gland.

That adrenaline may indeed act directly and rapidly on the pituitary gland to stimulate the secretion of ACTH was elegantly demonstrated by McDermott, Fry, Brobeck and Long (1950) who transplanted anterior pituitary tissue to the anterior chamber of the eye of hypophysectomized rats. When adrenaline was

had been stimulated by adrenaline in these isolated transplants. There is therefore little doubt that adrenaline influences the secretion of adrenal steroids by a pathway mediated by the anterior pituitary gland. Although adrenaline can act directly on the anterior pituitary gland it seems possible from the experiments of Brobeck, Long and their colleagues that adrenaline or noradrenaline may also act on hypothalamic centres.

THE SECRETION OF ACTH DURING STRESS

Chiefly from the work of Selye the importance of the secretion of adrenal hormones during a condition of stress and adaptation

has been recognized. The evidence accumulated by Long and his colleagues, and by G. W. Harris and others, which has been briefly mentioned above, supports the view that in stress and adaptation the secretion of ACTH by the anterior pituitary lobe results from hypothalamic stimuli together with the direct action of adrenaline or noradrenaline on the pituitary gland. Bush (1951) has recently made the interesting suggestion that there is a difference between the nature of the steroids secreted by the adrenal gland during a 'steady-state' and those liberated under conditions of stress. He suggests that in the steady-state condition there may be secreted factors which are present in the amorphous fraction of adrenal extracts, while under conditions of stress the secretion of 17-hydroxycorticosterone and corticosterone occurs in relatively large amounts. This interesting idea needs further experimental elucidation.

THE NATURAL ORIGIN OF ADRENAL STEROIDS

Sayers, Sayers, Fry, White and Long (1944) and Sayers, Sayers, Liang and Long (1946) found that the administration of ACTH to the rat and to the guinea-pig results in a rapid fall in the amount of cholesterol in the adrenal glands. More recently Conn, Vogel, Louis and Fajans (1950) have compared the effects of ACTH given to normal people and to patients with Addison's disease. In normal people there was a large increase in the urinary excretion of 17-ketosteroids and 11-oxysteroids, with a sharp diminution of serum cholesterol, primarily in the ester fraction. These changes were not seen in the patients with Addison's disease, and agree with the view that the adrenal steroids secreted by the gland find their origin in cholesterol. Isotope studies by Bloch and others support the idea that the formation from cholesterol of steroid hormones and their degradation products can occur in the body, and there seems to be no reasonable doubt that adrenal steroids are indeed formed from cholesterol, either from that already in the gland or from that which is deposited in the gland from the blood.

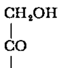
Recently Pincus and others have demonstrated the ability of the perfused bovine adrenal gland to secrete adrenal steroids in large amounts, and particularly to introduce an 11-oxy group

into perfused substances. There is now good evidence that in the adrenal gland there is present an enzyme system which is capable of introducing an 11-oxygen atom into the steroid nucleus.

METABOLISM OF ADRENAL STEROIDS

As was clearly demonstrated by Vogt (1943; 1950) the rate at which adrenal steroids are secreted by the adrenal glands and are inactivated by the peripheral tissues is very substantial. But the fate of the material which is thus altered is still not clear. Pregnanediol glucuronide has been isolated in relatively large amounts after the administration of deoxycorticosterone to normal men, while the fact that there is a substantial increase in the amount of urinary 17-ketosteroids when cortisone is given to patients with Addison's disease suggests that 17-ketosteroids are also metabolic products of cortisone.

The amount of reducing steroids, that is steroids with the side chain



at position 17, excreted in the urine by man is normally relatively small—about 0·3–1·0 mg/day, with a mean value of 0·5 mg/day, as estimated by the present methods. As Marrian (1951) has recently pointed out, the available methods for the estimation of adrenal steroids in urine may be extremely unsound, and the amounts of the excreted steroids may in fact exceed by many times those at present measured. With the techniques now available it can be shown that the amount of reducing steroids rises, perhaps tenfold or more, under conditions of stress, as for instance in diabetic ketosis (McArthur, Sprague and Mason, 1950).

ADRENAL STEROIDS AND METABOLIC PROCESSES

The changes in metabolism which follow experimental adrenalectomy are comparable in many ways with those seen in Addison's disease, and include a fall in the fasting blood-sugar level,

has been recognized. The evidence accumulated by Long and his colleagues, and by G. W. Harris and others, which has been briefly mentioned above, supports the view that in stress and adaptation the secretion of ACTH by the anterior pituitary lobe results from hypothalamic stimuli together with the direct action of adrenaline or noradrenaline on the pituitary gland. Bush (1951) has recently made the interesting suggestion that there is a difference between the nature of the steroids secreted by the adrenal gland during a 'steady-state' and those liberated under conditions of stress. He suggests that in the steady-state condition there may be secreted factors which are present in the amorphous fraction of adrenal extracts, while under conditions of stress the secretion of 17-hydroxycorticosterone and corticosterone occurs in relatively large amounts. This interesting idea needs further experimental elucidation.

THE NATURAL ORIGIN OF ADRENAL STEROIDS

Sayers, Sayers, Fry, White and Long (1944) and Sayers, Sayers, Liang and Long (1946) found that the administration of ACTH to the rat and to the guinea-pig results in a rapid fall in the amount of cholesterol in the adrenal glands. More recently Conn, Vogel, Louis and Fajans (1950) have compared the effects of ACTH given to normal people and to patients with Addison's disease. In normal people there was a large increase in the urinary excretion of 17-ketosteroids and 11-oxysteroids, with a sharp diminution of serum cholesterol, primarily in the ester fraction. These changes were not seen in the patients with Addison's disease, and agree with the view that the adrenal steroids secreted by the gland find their origin in cholesterol. Isotope studies by Bloch and others support the idea that the formation from cholesterol of steroid hormones and their degradation products can occur in the body, and there seems to be no reasonable doubt that adrenal steroids are indeed formed from cholesterol, either from that already in the gland or from that which is deposited in the gland from the blood.

Recently Pincus and others have demonstrated the ability of the perfused bovine adrenal gland to secrete adrenal steroids in large amounts, and particularly to introduce an 11-oxy group

TABLE 1. Physiological Activities of Adrenal Steroids

5 = high activity : 1 = slight activity

	Life Maintenance (rat)	Renal Function (dog)	Electrolyte Balance (rat, dog)	EV* Test (rat)	Muscle Work Test (Ingle) (rat)	Liver Glycogen (rat)	Diabetogenic Test (rat)
Corticosterone	1—2	2	1—2	1—2	3	3	3—4
17-Hydroxycorticosterone	2	1	?	2	5	5	5
11-Dehydrocorticosterone	1—2	2	1—2	?	2	2	3—4
Cortisone	2	1	?	2	4	4	5
11-Deoxycorticosterone	4	4	5	4	1	1	1
17-Hydroxy-11-deoxycorticosterone	3	4	?	2	1	1	1
Amorphous fraction	5	5	?	3	1—2	1	Nil ?

* Everse de Fremery work test

which may cause fatal hypoglycaemia, and a disturbance of the electrolyte balance, particularly a loss of sodium and chloride ions from the body, with retention of those of potassium. There is a loss of fat, a subnormal breakdown of protein, with a consequent tendency to retain protein in the body, and a particularly rapid rate of oxidation of carbohydrate and loss of glycogen during fasting; the fasting respiratory quotient may be unusually high. There occurs an associated muscular weakness and a diminished resistance to many forms of stress. Untreated Addison's disease sometimes, and experimental adrenalectomy usually, is fatal, although it has been observed that pregnancy or pseudo-pregnancy may surprisingly prolong the time of survival after experimental adrenalectomy. The activity of progesterone as an adrenal hormone may here be of significance though probably this is not the only factor concerned.

Conversely the metabolic effects which follow the administration of crude adrenal extracts or adrenal steroids, and of ACTH, as well as the changes in metabolism observed in Cushing's syndrome, are largely the converse of those which are seen in Addison's disease and in adrenalectomized animals. Diabetes is often found, although sometimes nothing more than a diminished sugar tolerance is seen. There is a tendency to retain sodium chloride, with a fall of blood potassium level which may involve serious cardiac damage. An excessive deposition of fat is sometimes accompanied by a loss of protein from the body, and a high liver glycogen content with signs of increased glyconeogenesis from protein. Growth may be diminished or prevented in young animals. The virilizing syndrome seen in humans as the result of hyperfunctional tumours of the adrenal cortex is probably associated with an abnormally large production of androgens by the adrenal cortex, and we shall not discuss this condition here.

In the experimental administration of separated adrenal steroids interesting differences between the activities of the individual substances are observed. Some of these are set out in Table 1. Here the 'renal function test' in adrenalectomized dogs is concerned with the prevention of the symptoms of water intoxication which may follow the administration of a water

1951; cf. Ingle, 1950). It is true, however, that in Addison's disease cortisone alone will not always prevent an excessive loss of sodium chloride, in the doses that have been administered as yet (Sprague, Mason and Power, 1951), although it is possible that with larger doses such an effect may be seen.

Slessor (1951) finds that the plasma of fasting patients with Addison's disease contains an abnormally large amount of antidiuretic substance, possibly of posterior-pituitary origin. In Addison's disease the response to a water load is much delayed, but treatment with cortisone can restore the time and degree of diuresis to normal. Slessor puts forward the interesting hypothesis that adrenal steroids of the cortisone type may be concerned with the inactivation of circulating posterior pituitary hormones.

INFLUENCE OF ACTH ADRENAL STEROIDS ON CARBOHYDRATE METABOLISM

Administration of cortisone or ACTH to rats under suitable conditions is followed by hyperglycaemia and glycosuria In Cushing's syndrome there is generally observed an impairment of sugar tolerance or frank diabetes, and from the urine of a case of diabetes with Cushing's syndrome Mason and Sprague (1948) isolated 17-hydroxycorticosterone in such amount (about 8 mg/day) that it appeared to account for about one-half of the total excretion of adrenal steroid Sprague, Mason and Power (1951) have observed an exacerbation of diabetes in coexisting Addison's disease and diabetes when as little as 30 mg/day of cortisone was given for a period of 5 days. This exacerbation of diabetes was associated with a rise in the rate of excretion of both nitrogen and water When cortisone was given to normal humans, on the other hand, no rise in fasting blood sugar was observed. In four out of twelve cases there was a slight impairment of carbohydrate tolerance, but in two of these four cases there was a family history of diabetes. In many of these cases the repeated dose of cortisone exceeded 100 mg/day.

In contrast with these observations on cortisone, diabetes has been found to follow the administration of large doses of ACTH to a significant proportion of human beings, in particular by Conn, Louis and Johnston (1949). This difference between the

load, while the test for 'electrolyte balance' depends on the ability of the steroids to induce potassium excretion and sodium and chloride retention in normal or adrenal-deficient animals. The 'Everse de Fremery' (EV) work test depends on the extent of recovery of the muscle of the rat from a brief period of electrical stimulation, while the 'Ingle work test' in the rat is concerned with a more prolonged effect. In this test the gastrocnemius muscle of an anaesthetized rat is weighted with 100 gm and then stimulated to contract three times a second for 48 hours or until death. The effectiveness of hormones in the muscle work test of Ingle is quite different from that in the Everse de Fremery work test, as will be seen from the table. The 'liver glycogen test', first introduced by Long, Katzin and Fry (1940), depends on the ability of an injected substance to induce an increase in the amount of glycogen in the liver of the fasting normal or hypophysectomized rat.

As a result of classifications of the type set out in this table statements which may be dangerous and extremely misleading are sometimes made about the pharmacology of adrenal steroids. Since the observed activity of the steroids often varies not only from species to species but also with different conditions in the same animal, it is most undesirable that generalizations be made which involve extrapolations from one species to another, or even from one condition to another. It can be tentatively said in general terms that the presence of an 11-oxygen or hydroxyl group in the steroid nucleus depresses the activity of the steroids in the electrolyte balance and renal function tests and enhances effects on carbohydrate metabolism and on muscle work, while the presence of a 17-hydroxyl group depresses electrolyte activity and enhances influence on carbohydrate metabolism. In some early work of Thorn and his collaborators it was found that 17-hydroxycorticosterone and cortisone both caused a rise in the excretion of sodium and chlorine in short-term experiments in rats or dogs. With more prolonged treatment these substances nevertheless usually induce a retention of sodium and chloride (Thorn, Forsham, Bennett, Roche, Reiss, Slessor, Flink and Sommerville, 1949; Fourman, Bartter, Albright, Dempsey, Carroll and Alexander, 1950; Sprague, Mason and Power,

in cats by the administration of ACTH, although adults of this species are very sensitive to the diabetogenic action of growth hormone. Any influence of growth hormone on purine metabolism awaits further investigation.

It is of interest to note that when gluconeogenesis occurs under the influence of adrenal steroids it is sometimes impossible to account for the extra carbohydrate appearing on the basis of the classical figures for the conversion of protein to carbohydrate.

INFLUENCE OF ADRENAL STEROIDS AND ACTH ON PROTEIN METABOLISM

Growth is inhibited by large doses of ACTH or adrenal steroids such as cortisone, although when immature adrenalectomized rats are treated with small doses of cortisone they are able to survive and grow at a subnormal rate. In general the administration of cortisone and allied substances results in protein catabolism or failure in the anabolism of protein, this leading to the excretion of nitrogen in excess of the intake, that is to a negative nitrogen balance, and thus to a reduction in the total amount of protein in the body. Similar observations have been made in Cushing's syndrome. Treatment with cortisone also leads to an inhibition of the growth of hair, and a depression in the rate of granulation of wound tissue, both of which may be associated with a diminished protein anabolism or increased protein catabolism. Cortisone can also inhibit the growth of tumour tissues under certain conditions.

In the human being the alleviating action of cortisone on rheumatoid arthritis is not necessarily dependent on protein breakdown, at least as judged by the gross nitrogen balance. Sprague, Mason and Power (1951) found that towards the end of a period of administration of 200 mg/day of 17-hydroxycorticosterone only minor changes in nitrogen balance, urinary excretion of creatine nitrogen and uric acid had occurred in a patient with rheumatoid arthritis, although unmistakable evidences of anti-rheumatic activity were observed Approximately the same metabolic and clinical results were observed in this case with a similar dose of cortisone.

All the 11-oxygenated steroids of the adrenal cortex cause

effects of ACTH and of cortisone in the human may indicate that the secretions of the adrenal cortex are more potent in influencing carbohydrate metabolism than is cortisone. Unfortunately, the action of 17-hydroxycorticosterone on carbohydrate metabolism has not yet been adequately investigated. It should be noted that the diabetes induced by ACTH, and also that produced in animals by treatment with adrenal steroids, is associated with a striking resistance to the action of insulin.

Conn and his colleagues have found that in the diabetes produced in human beings by ACTH the blood glutathione content falls in a significant fashion, and conversely that the development of diabetes may be prevented, or the condition alleviated, by the administration of reduced glutathione. A rise in the excretion of uric acid occurs in human beings under treatment with ACTH and Conn has considered the possibility that the diabetes which occurs may be associated with the metabolic upheaval of purine and glutathione metabolism thus induced. More recently he has compared the influence of ACTH on the carbohydrate and purine metabolism of Dalmatian and mongrel dogs (Fajans, Conn, Johnson and Christman, 1951). In all the dogs the excretion of total purine nitrogen (that is uric acid + allantoin) was raised by about 25 per cent as the result of the administration of ACTH. In the Dalmatian dogs uric acid was responsible for the major increment in the total purine nitrogen, while in the mongrels it was predominantly extra allantoin that was excreted. Treatment with ACTH resulted in a typical diabetic glucose tolerance curve in two Dalmatian dogs and in one mongrel dog, but not in two other mongrel dogs, and Conn concludes that there is a relationship between the diabetic response and the rise in uric acid excretion. Loubatières and Bouyard (1951) have recently emphasized the possibility that the development of diabetes may sometimes be associated with the presence in the blood of significant amounts of alloxan. If indeed alloxan is a normal intermediary in purine metabolism it may be that alloxan or a closely related substance should be seriously considered as a possible causative factor in ACTH diabetes, and thus in adrenal steroid diabetes In our own laboratory E. Reid (1951) has been quite unable to produce diabetes

the response of cells to a variety of noxious stimuli. These influ-
[illegible]

steroids in bringing about these changes is still a mystery.

ACTH AND ADRENAL STEROIDS AND RHEUMATOID ARTHRITIS

The possibility that tissue permeability might be a factor of some importance in the aetiology of rheumatoid arthritis has been considered by many workers, and since the mucopolysaccharide hyaluronic acid is known to be a substance of considerable importance in connective tissue the possible influence of adrenal cortical hormones on the enzyme hyaluronidase, which splits hyaluronic acid and can thus increase tissue permeability, was of some interest. In the blood serum of both animals and man there is a non-specific inhibitor of hyaluronidase which can depress the activity of hyaluronidase from any source. The amount of this inhibitor in the blood rises in many infectious diseases and a significant elevation is observed in rheumatic conditions. In general the elevated values in patients with rheumatoid arthritis and rheumatic fever fall during treatment with cortisone and ACTH (Hakanson and Luft, 1949; Schmith and Faber, 1949; Dorfman and Moses, 1950; Adams, Kelley, Dwan and Glick, 1951; Benditt, 1951).

Experiments carried out by Opsahl (1949, 1951) at Yale have provided evidence that adrenal steroids, but not deoxycorticosterone acetate, are themselves capable of inhibiting the action of hyaluronidase in promoting the spread of dyestuffs administered intradermally to animals. Whether this effect is on the enzyme or on the substrate is not known for certain. but Opsahl has produced some evidence that the activity of hyaluronidase *in vitro* is depressed in the presence of adrenal extracts. Though the results of such experiments must be interpreted with great caution it may be that in the *in vivo* experiments the primary influence of adrenal steroids is on the enzyme and not on its substrate. The significance of these results with respect to rheumatoid arthritis is not at all clear, and perhaps

atrophy of the thymus of the rat when administered in high doses, and there are similar changes in lymph nodes and in the spleen. This loss of lymphatic tissue is related to and perhaps dependent upon an increased destruction of lymphocytes, since the mitotic activity of the atrophic glands is undiminished.

How far the influence of adrenal steroids and ACTH on inflammation and the phenomena of immunity should be considered here is doubtful, but it may be noted that the response of tissues to many different types of injurious influence may be altered by the administration of ACTH or adrenal steroids. For example, the acute inflammatory reactions in the eye to a wide range of agents may be inhibited under the influence of adrenal cortical stimulation or local treatment with adrenal steroids. Although there are some experimental results to the contrary the consensus of opinion appears to be that ACTH has no substantial effect on the development of antibodies, and although some hypersensitive phenomena are altered by adrenal cortical stimulation and ACTH, others are not.

THE INFLUENCE OF ACTH AND ADRENAL STEROIDS ON FAT METABOLISM

Adrenalectomy in the rat results in a diminution in the amount of stored fat (Simpson, Dennison and Korenchevsky, 1934), and a fall in the total fat is associated with an increase in the protein content of the body (Schiffer and Wertheimer, 1947). Conversely the administration of corticosterone or of cortisone to animals results in an increase in the total body fat, particularly in the liver fat, while a fatty liver is often seen in Cushing's syndrome. Ketonuria has been observed to follow the administration of ACTH and of cortisone, both to patients and to animals. One may conclude that adrenal steroids depress the metabolism of fat as well as that of carbohydrate, but whether this is a direct or indirect effect is still uncertain.

The general conclusion to be drawn from the study of the influence of ACTH and adrenal steroids on organic metabolism is that the oxidation of carbohydrate and fat is depressed while protein anabolism is inhibited or protein catabolism enhanced. The latter phenomena probably play some part in diminishing

REFERENCES

ADAMS, F. H., KELLEY, V. C., DWAN, P. F., and GLICK, D. (1951). *Proc. Second Clinical* ACTH *Conf*, ed J R. Mote, vol. ii, p. 572. Blakiston Co. N.Y.

BENDITT, E. B. (1951). *Proc. Second Clinical* ACTH *Conf.*, ed. J. R. Mote, vol. ii, p 585. Blakiston Co N.Y.

BUSH, I. E. (1951). *J. Endocr.* **7**, lxxxiii.

CONN, J. W., LOUIS, L H., and JOHNSTON, M. W. (1949). *J. Lab. clin. Med.* **34**, 255.

— VOGEL, W. C., LOUIS, L H., and FAJANS, S. S. (1950). *J. Lab clin. Med.* **35**, 504.

DIXON, H B. F., MOORE, S, STACK-DUNNE, M. P., and YOUNG, F. G (1951). *Nature, Lond.* **168**

— STACK-DUNNE, M. P., YOUNG, F G, and CATER, D. P. (1951). *Nature, Lond* **168**.

DORFMAN, A, and MOSES, F E. (1950). *Proc. Soc. exp. Biol., N Y.* **73**, 667.

DOUGHERTY, T F, and WHITE, A. (1944). *Endocrinology*, **35**, 1.

FAJANS, S. S., CONN, J. W., JOHNSTON, D V., and CHRISTMAN, A A. (1951). *Endocrinology*, **49**, 225

FOURMAN, P., BARTTER, F. C, ALBRIGHT, F., DEMPSEY, E., CARROLL, E, and ALEXANDER, J. (1950) *J. clin Invest* **29**, 1462

GROOT, J. de, and HARRIS, G W (1950). *J Physiol.* **111**, 335.

HAKANSON, E J, and LUFT, R. (1949) *Acta Endocr.* **3**, 318

HARRIS, G. W (1951) *Brit med J* **2**, 559

HECHTER, O, JACOBSEN, R P, JEANLOZ, R, LEVY, H, PINCUS, G., and SCHENKER, V (1950). *Proc Amer. Diabetes Ass.* **10**, 39

— ZAFFARONI, A., JACOBSEN, R P, LEVY, H, JEANLOZ, R W., SCHENKER, V, and PINCUS, G (1951) *Recent Prog Hormone Res.* **6**, 215

HILLS, A G, FORSHAM, P. H, and FINCH, C. A. (1948) *Blood*, **3**, 755

HUME, D. M, and WITTENSTEIN, G J (1950) *Proc First Clinical* ACTH *Conf*, ed J R. Mote, p 134 Blakiston Co N.Y

INGLE, D J. (1950) *J. clin. Endocr* **10**, 1312

[illegible]

1251, *C R Soc. de Biol, Paris*, **145**, 493

McARTHUR, J W, SPRAGUE, R. G, and MASON, H L (1950) *J Clin. Endocr.* **10**, 307.

McDERMOTT, W V, FRY, E G, BROBECK, J R, and LONG, C N H (1950). *Yale J Biol Med* **23**, 52, *Proc Soc exp Biol, N Y* **73**, 609

MARRIAN, G F (1951) *J Endocr* **7**, lxix

MASON, H L (1950). *J biol Chem* **182**, 131

— and SPRAGUE, R G (1948) *J biol Chem* **175**, 451

will not be clear until we know much more about the chemistry of connective tissue.

Because the development of jaundice sometimes affords marked relief to patients with rheumatoid arthritis, Sprague, Mason and Power (1951) have investigated the adrenal steroid metabolism of jaundiced patients, and have found that adrenal steroids are excreted at a normal rate. There is therefore no reason to believe that the relief of the symptoms of rheumatoid arthritis in these patients results from an increased rate of secretion of steroids by the adrenal glands.

CONCLUSION

The only general conclusion that can be drawn at the present time is that the adrenal steroids, particularly 11-oxy steroids, are capable of reducing the response of the tissues to the presence of some types of toxin. The mechanism for this reduction in response is far from clear although it could conceivably be associated with the inhibition of protein anabolism or the stimulation of protein catabolism that may be observed in such instances. Despite the magnitude of the research that has been carried out in the three years which have elapsed since Hench's fundamental observations we are still very far from understanding the mechanisms which underlie them There is no evidence that rheumatoid arthritis or other conditions which are favourably affected by adrenal steroids are associated with a deficiency of adrenal hormones. Indeed there is growing evidence that such is not the case. It therefore appears that a state of hormonal excess induces a favourable influence on diseases which are not characterized by adrenal-cortical insufficiency. It is clear that the knowledge which has accumulated as the result of the use of ACTH and cortisone both experimentally and clinically, is striking at the heart of many long-held views.

III

Recent Light on Mammary Function

H. D. KAY

If a lecture with this title had been given just before the war, it would perhaps have been possible in a single hour to outline most of the advances in knowledge that had been secured during the previous decade. But even to outline the important advances in this field since 1940 is quite impossible in a single lecture. I shall not attempt it, but instead I shall talk mainly about two topics in which my colleagues at Shinfield and I have been particularly interested for several years, first, the synthesis of milk in the mammary gland, second, the phenomenon of 'let-down' or 'draught'.

As mammals ourselves, it is remarkable that even now we know so little about the astounding function that distinguishes the zoological class *mammalia* from other creatures, a function which has undoubtedly helped that class to its present outstanding position amongst living animals. Nature is full of major astonishments, but this provision by the mammalian mother out of her own tissues of a continuous and beautifully balanced commissariat to meet the first needs of the next generation is not only amazing in itself but like other cardinal evolutionary steps still lacks an adequate explanation as to its origin and early development.

Both quantitatively as regards volume of output, and qualitatively as regards the character of that output, the activity of the functional mammary gland is surprising. Quantitatively, for example, the average good cow, in a single lactation, may produce as much as four times her own dry weight of milk solids, and many cows produce far more than that. To put it slightly

NELSON, D. H., SAMUELS, L. T., WILLARDSON, D. G., and TYLER, F. H. (1951). *J. clin. Endocr.* **11**, 1021.
OPSAHL, J. C. (1949). *Yale J. Biol. Med.* **21**, 255, 433, 487.
— (1951). *Trans. Second Conf. Adrenal Cortex*, Josiah Macy Jr. Foundation, ed. E. P. Ralli, p. 115.
PICKFORD, M., and VOGT, M. (1951). *J. Physiol.* **112**, 133.
REID, E. (1951). *J. Endocr.* **7**, xxxvi.
REID, J., WATSON, R. D., COCHRAN, J. B., and SPROULL, D. H. (1951) *Brit. med. J.* **2**, 321.
REICH, H., NELSON, D. H., and ZAFFARONI, A. (1950). *J. biol. Chem.* **187**, 411.
REINHARDT, W. O., and LI, C. H. (1951). *Proc. Soc. exp. Biol., N.Y.* **77**, 229.
SAYERS, G. (1950). *Physiol. Rev.* **30**, 241
— SAYERS, M. A., FRY, E. G., WHITE, A., and LONG, C. N. H. (1944). *Yale J. Biol. Med.* **16**, 361.
— SAYERS, M. A., LIANG, TSAN-YING, and LONG, C. N. H. (1946). *Endocrinology*, **38**, 1.
— SAYERS, M. A., and WOODBURY, L. A. (1948). *Endocrinology*, **42**, 379.
— WHITE, A., and LONG, C. N. H. (1943). *J. biol. Chem.* **149**, 425.
SCHIFFER, F., and WERTHEIMER, E. (1947). *J. Endocr.* **5**, 147.
SCHMITH, K., and FABER, V. (1949). *Acta Endocr.* **3**, 310.
SIMPSON, S. L., DENNISON, M., and KORENCHEVSKY, V. (1934). *J. Path. Bact.* **39**, 569.
SLESSOR, A. (1951). *J. clin. Endocr.* **11**, 700
SMITH, P. (1927) *J. Amer. med. Ass.* **88**, 158.
— (1930). *Amer. J. Anat.* **45**, 205.
SPEIRS, R. S., and MEYER, R. K. (1949) *Endocrinology*, **45**, 403.
SPRAGUE, R. G., MASON, H. L., and POWER, M. H. (1951). *Recent Prog. Hormone Res.* **6**, 315.
STACK-DUNNE, M. P., and YOUNG, F. G. (1951). *J. Endocr.* **7**, lxvi.
THORN, G. W., FORSHAM, P. H., BENNETT, L. L., ROCHE, M., REISS, R. S., SLESSOR, A., FLINK, E. B., and SOMMERVILLE, W. (1949). *Trans. Ass. Amer. Physicians*, **62**, 233
VOGT, M. (1943) *J. Physiol.* **102**, 341
— (1944) *J. Physiol.* **103**, 317.
— (1950). *Brit. med. J.* **2**, 1242.
— (1951) *J. Physiol.* **113**, 129
YOUNG, F. G. (1951) *Brit. med. J.* **2**

mammary circulation of some 200 litres per hour. As regards milk production in proportion to her weight, the efficient human mother is not far behind the cow, but the domestic sow with, say, a litter of twelve dependent on her efforts beats them both to a frazzle!

Synthesis of the major milk constituents, casein, lactose and milk fat, which are different in chemical composition from anything brought to the mammary gland by the circulating blood, is, at or near the broad peak of the lactation period, a very rapid process.

A highly diagrammatic picture of one of the group of cells whose astounding activities some of us have been endeavouring to unravel is shown in Fig. 1.

The points I wish to make at this stage are (1) that milk secretion is not only a rapid but a rhythmic process, (2) that all the major constituents of milk are synthesized in one and the same cell, (3) that the cells do not undergo destruction during or after the synthetic process, but that the identical cells appear to remain functional for at least a large part of the lactation period.

SYNTHESIS IN THE MAMMARY GLAND

First, a brief word about methods of investigation.

1. *The A-V method* The obvious method, and one that has been much used in recent years, despite its drawbacks, is to determine the level of possible precursors in the arterial blood arriving at the active gland, and the level of the same constituents in the venous blood leaving the gland, usually called the A-V method. A sample of arterial blood from any major artery will do, the mammary venous blood can be taken from any of the large veins leaving the gland (since there is considerable anastomosis in the venous system of the gland). This method, employed mainly on the cow and the goat, has given useful information (see Table 1). The anaesthetized animal has sometimes been used; this avoids certain troubles but introduces others. There are serious drawbacks, both practical and theoretical, to the A-V method, whether the animal is anaesthetized or not.

2. *Perfusion of the isolated surviving mammary gland.* This also has

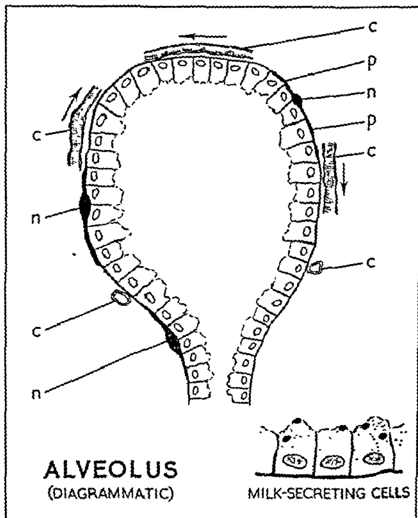


FIG. 1 *c* indicates capillary blood vessel
n indicates nucleus of myoepithelial cell
p indicates process of myoepithelial cell

differently, in one day a reasonably efficient udder produces *milk solids equivalent to half its own dry weight. For every litre* of milk the cow produces, something like 400 litres of blood must go through the gland, so that an average good cow has a

rapidly growing calf or goatling. From what constituent or constituents of the circulating blood does the mammary gland make casein—which, it will be remembered, is a phosphoprotein?

It has been known for some time, as the result of the several investigations, that the amino-acid content of blood diminishes in traversing the lactating mammary gland in the cow. But this fall in the plasma concentration of amino-acids is not very great, and does not seem sufficient to provide all the nitrogen required by the active gland for protein synthesis. An obvious source might be circulating blood proteins, which, by partial fragmentation or even complete breakdown to amino-acids and resynthesis in the alveolar cells, might be given the characteristic amino-acid composition of casein or lactalbumin. Reineke, Williamson and Turner (1941) reported an appreciable fall in the content of one of the plasma glycoproteins as the blood passed through the udder, a finding which gave some support to this suggestion.

As regards the phosphorus of casein, which is mainly combined as orthophosphate esterified with the hydroxy group of serine and possibly other hydroxy-amino-acids in the casein molecule, there was previously some good, but not entirely conclusive evidence that it came from the inorganic phosphate fraction of the circulating blood and not from the combined phosphate in the blood. Careful work had indeed shown that, whilst there was a substantial diminution of inorganic phosphate in the blood traversing the active gland, there was little, if any, change in the other phosphorus compounds.

Quite recent work (1951) of J. M. Barry, a former Shinfield colleague, who has kindly allowed me to quote results which are not yet published, has shown that the lactating mammary gland of the goat or the cow takes free lysine and tyrosine from the blood stream to provide all the lysine and tyrosine of casein, and takes inorganic phosphate to provide all the casein phosphorus, no significant amounts of plasma protein lysine and tyrosine are used for casein synthesis.

These findings rest on the use of labelled tyrosine and lysine, containing radioactive C^{14} in the carboxyl group, and labelled inorganic phosphate containing P^{32}. These were in one experi-

given valuable results, but there are, again, practical difficulties, in this case concerned with blood aeration, prevention of clotting, maintenance of normality of composition of blood from which constituents are continually being removed and, for the most part, not being replaced, possible loss of hormones from the blood, deterioration and oedema of the gland, etc.

TABLE 1. Blood (Plasma) Constituents that Diminish in Concentration in Passing through the Mammary Gland

Oxygen	Amino-acids
—	Glycoprotein
Glucose	—
Acetate	Inorganic phosphate
β-hydroxybutyric acid	Calcium
Neutral fat	*Potassium

* Inference only

[illegible] suspended in oxygen-

when working under more strictly physiological conditions.

4. *Use of isotopes.* This has given quite outstandingly valuable findings in the last four or five years, and much of the following account is concerned with findings so obtained.

PROTEIN SYNTHESIS

Casein and lactalbumin are the characteristic proteins of milk. They do not occur elsewhere in the tissues or secretions. In cow's milk there is some 2·8 to 3·0 per cent of casein, that is, about one-quarter of the total solids of the milk. The proportion in goat's milk is similar but there is less in human milk; the more slowly growing baby requires less protein per unit time than the

The second observation was a general one arising from the findings of several groups of workers, namely, that acetate probably played a central rôle in fat metabolism. Rittenberg and Block (1945), using C^{13} for labelling, had shown that mice and rats could synthesize, from acetate, body fats containing fatty acids of various lengths of chain involving the multiple condensation of C_2 units.

The third observation, made at Shinfield by Folley and French (1950), showed that mammary gland slices (from lactating ruminants) respiring in Warburg manometers would utilize acetate and give a high respiratory quotient in doing so, i.e. that acetate was probably being synthesized into fat by such slices. Slices of non-ruminant mammary glands appeared unable to utilize acetate, though they gave a high RQ when respiring in glucose solutions.

Quite recently, Popják, French and Folley (1950 and 1951), by administering radioactive sodium acetate (with the C^{14} in the COOH group) intravenously into a lactating goat, have shown that the lower fatty acids of the milk fat subsequently secreted contained radioactive carbon in very appreciable quantities. A proportion of the circulating acetate was quite rapidly utilized for the synthesis of these fatty acids The plasma fatty acids contained much less C^{14} than the milk fatty acids. The steam-volatile fatty acids from the milk fats secreted in the first stages had a much higher specific activity than the longer-chain, non-volatile acids, indicating their independent synthesis. There seemed no doubt that the hypothesis, which had previously held the field, that the short-chain acids originate from the degradation of the long-chain acids, particularly of oleic, would have to be seriously modified.

On further examination of the fatty acids of the milk fat formed subsequently to the injection, it was observed that the specific activity of the capric (decanoic) acid was greatest in the fat secreted during the first twelve hours, but that in the fat secreted subsequently, the specific activity was greater in the longer-chain acids but not in stearic nor oleic, which appear to have a different origin from that of the other fatty acids. Stearic and oleic acids might be derived from the neutral fat of the blood

ment injected into the blood stream of a lactating goat, and in a second, added to the blood perfusing a surviving excised mammary gland of a cow. Specific activities were measured in various components of blood and milk taken at intervals. Quite similar results were obtained, using either of the two methods.

From these experiments, it seems probable that the casein of milk is synthesized in the alveolar cells from free amino-acids in the blood, though they do not completely rule out the possibility that a portion of the casein molecule not tyrosine nor lysine might come from some other nitrogenous constituent of the blood of larger molecular size than a free amino-acid. The experiments done so far tell us nothing about lactalbumin synthesis; this protein may well come from some more complex source than the circulating free amino-acids.

FAT SYNTHESIS

The milk fat of most mammals, and especially of ruminants, is characterized by having an appreciable percentage of short-chain fatty acids. Early experiments with unsatisfactory technique showed, apparently, that the phospholipin of the blood was the source of milk fat. Later (Graham, Jones and Kay, 1936), we showed at Shinfield, using the A-V method, that there was no change in phospholipin, but a definite though small diminution in neutral fat—i.e. in the plasma triglycerides, which do *not* contain short-chain fatty acids. This material might, of course, be oxidized by the gland for energy purposes, and not be a source of milk fat at all.

Three interesting series of observations were made in the late thirties and forties, the first, by Barcroft and his collaborators (see Elsden and Phillipson, 1948), was that much of the carbohydrate of the diet of the ruminant was fermented in the rumen to acetic, propionic, and butyric acids, particularly the first of these. These acids were taken up rapidly, apparently through the rumenal wall, into the circulating blood. Later, it was shown, by McClymont (1949), that ruminant blood contained as much as 10–12 mgm per cent of acetate which diminished markedly in amount in passing through the lactating mammary gland.

radioactive. Thus, it is quite clear that the cells of bovine mammary tissue can synthesize from acetate at least one milk constituent—fat (Cowie *et al.*, 1951).

An interesting difference between the mammary tissue of ruminants and non-ruminants has been observed. Mammary gland slices of ruminants, e.g. sheep or goat, will utilize acetate to synthesize fat, but mammary gland slices of non-ruminants, e g. the rat, will utilize acetate only in presence of glucose and not in its absence. This finding links up with the fact that in ruminant blood the normal amount of glucose is much lower and that of acetate higher than in the non-ruminant. The blood 'sugar' of the ruminant as ordinarily determined is rarely above 0 06 per cent and a fair proportion of this—perhaps one-third or more—is not glucose. On the other hand, the blood 'sugar' of the non-ruminant is rarely below 0·1 per cent and often considerably above this figure, and is mainly genuine glucose. It is far from misleading to say that blood acetate tends to fulfil in the ruminant organism certain of the functions performed in the non-ruminant by glucose.

Another related difference has recently been found between the metabolic activity of the ruminant and the non-ruminant. In non-ruminant mammary gland slices, metabolizing a mixture of acetate + glucose, insulin added *in vitro* will increase both acetate utilization and glucose uptake—an interesting example, incidentally, of the direct action of insulin on isolated tissue. But ruminant gland slices from the lactating goat or sheep are inert to insulin both in acetate and in acetate + glucose.

By using radioactive acetate in the Warburg manometer, it has been possible to show directly that rat mammary gland slices will incorporate acetate carbon into fatty acids only in the presence of glucose, but not in its absence. The addition of insulin will cause a further incorporation of acetate carbon, the specific activity of the synthesized mixed fatty acids going up by 100 per cent. On the other hand, ruminant mammary gland slices (in this case from lactating ewes) will incorporate acetate carbon in absence of glucose, though the incorporation goes faster in presence of the sugar. In these experiments, the glycerol of the synthesized fat was found to be non-radioactive, indicating that

plasma, which fat, as I mentioned a moment ago, diminishes in quantity in going through the mammary gland, and which contains little, if any, lower fatty acid.

The butyric and also the caproic (hexanoic) acids synthesized were then chemically degraded in such a way that the radioactivity of each of the individual carbon atoms could be measured. Only the carbon atoms 1 (COOH), 3 and 5 were radioactive (Popják, French, Hunter and Martin, 1951).

From this work the conclusion has been drawn that all the milk fatty acids, up to and including palmitic acid, can be formed by the stepwise elongation of a shorter acid by the addition of a C_2 compound derived from acetate. Butyric acid is synthesized from two acetate molecules—carboxyl to methyl—but it is inferred, as a result of good experimental evidence that cannot be discussed here, that only about 40 per cent of the butyric acid is synthesized in this way, the other 60 per cent being derived from a non-isotopic C compound in the circulating blood, possibly β-hydroxybutyric acid, which American work (Shaw and Knodt, 1941) has previously shown to diminish in concentration in blood going through the mammary gland. Caproic acid is synthesized by the elongation of the butyric acid chain at the carboxyl end by the addition of acetate.

Though the evidence provided by these elegant experiments seems convincing that the synthesis, from acetate, of several of the milk fatty acids takes place on the spot, i.e. in the alveolar cells, it is remotely possible that the labelled carbon of the acetate might have been incorporated into more complex milk fat precursors in the liver or elsewhere and transported to the mammary gland An experiment was, therefore, carried out on the isolated perfused udder taken from a lactating cow, labelled sodium acetate being added to the fluid perfusing one half of the udder, and labelled sodium bicarbonate to that perfusing the other half. (I ought perhaps to mention that the two halves of the udder, as regards their circulatory and milk secretory systems, are anatomically separate.) The fatty acids of the milk fat produced in the acetate half were radioactive, the specific activities diminishing with the average chain length. None of the fatty acids of the milk fat produced in the bicarbonate half was

to a low level. It was shown experimentally at Shinfield that a ration composed of a high proportion of concentrates and a low proportion of hay regularly caused a marked fall in the percentage of fat in the milk. Using cows with rumenal fistulae, indications were obtained that such a ration does not stay long enough in the rumen to allow the normal proportion of its cellulose and other carbohydrates to be fermented into acetate. If, when the milk fat percentage is depressed, more roughage, hay or straw, is added to the diet, rumenal time appears to be increased, more acetate is produced and in most cases the milk fat percentage rises (Balch, 1951).

It seems clear that one of the necessities for ensuring a normal butter-fat percentage in the commercial cow is the maintenance of a sufficiently large fermentation rate and acetate uptake from the rumen. Information on how to attain these in dairy farming practice is steadily accumulating.

THE SYNTHESIS OF LACTOSE

The disaccharide, lactose (glucose-β-galactoside), occurs in nature, so far as I am aware, in milk only Galactose, of course, occurs in combination in several other substances of biological importance.

It has been known for some twenty years that, as blood traverses the active mammary gland, its glucose content diminishes considerably. In our own findings (Graham, Jones and Kay, 1936) as much as 20 per cent of the reducing sugar in bovine arterial blood was lost in passage through the udder. The uptake of glucose during a given time was sufficient, if it had been quantitatively transformed into lactose, to cover the whole of the lactose secreted. Of course, the glucose might have been used by the gland for other purposes, e.g. oxidized to provide for energy requirements, or transformed into fat.

I mentioned earlier that there was experimental evidence that one of the sugar-containing plasma globulins—a glycoprotein with a galactose-mannose moiety—also diminished in amount in going through the goat's mammary gland, though the amount removed in terms of galactose-mannose seemed too small to account for more than a small fraction of the lactose secreted.

mammary tissue, at least *in vitro*, cannot make use of acetate carbon for the synthesis of glycerol. Glycerol seems to be a limiting factor for fat synthesis in mammary gland slices; addition of glycerol to rat slices metabolizing glucose + acetate has a stimulating effect similar to that of insulin, and will also increase fat synthesis in sheep udder slices metabolizing in acetate. In the intact animal it has been shown recently that the glycerol combined in the milk fat is derived from circulating glucose.

It is interesting to speculate on the reason why, in most milk fats and particularly in that of the ruminant, the presence of fatty acids with short carbon chains is marked, whereas in most other fats and in animal depot fats, short-chain fatty acids are present in almost negligible amounts. It may well be that in the mammary cells with the tremendous rate of synthesis of fat in progress, some only partly-completed structures slip out of the alveolar factories. An alternative teleological suggestion that perhaps the short-chain acids are better utilized than the long-chain ones in the nutrition of the suckling seems to have no evidence in its favour. This emphasis on the stepwise synthesis of milk fat from acetate, or a 2-carbon atom compound, by no means rules out the possibility that milk fat *can* be made by the gland from other sources such as blood glucose. In fact, recent experiments of French and Popják (1951) with the lactating rabbit have shown that, if radioactive glucose is given, the fatty acids of the milk fat secreted contain radioactive carbon Some of my colleagues have quite recently found that lactating rat mammary gland slices will synthesize radioactive fatty acids from such glucose It seems highly probable that glucose is broken down to 2-carbon atom fragments before resynthesis into milk fat.

One application of the more recent knowledge of ruminant metabolism and milk synthesis may be mentioned. The fat percentage in cow's milk is of importance both in relation to the nutritional value of the milk, if consumed as a liquid, and also to manufacturers of milk products It is not unusual for a dairy farmer, who is feeding a high-concentrate–low-roughage ration, or whose animals are eating nothing but the succulent young grass of a modern rapidly-growing ley, to find that the percentage of fat in the mixed milk of his herd falls precipitately

intravenously into a lactating goat, and blood and milk samples were taken at intervals after injection. The specific activities of blood glucose, of lactose, and of the glucose and galactose residues in each sample of lactose, were measured. Those of the glucose and galactose derived from the hydrolysed lactose were equal. In a second experiment of Barry's, glucose labelled with C^{14} in position 1 was injected into a lactating goat (see Fig. 2).

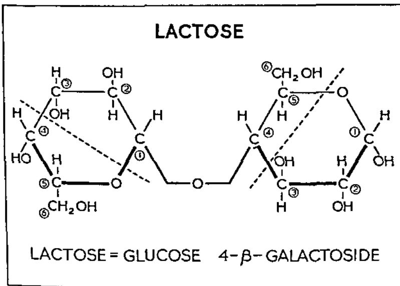


FIG. 2. Formula for lactose, illustrating Barry's findings on the mechanism of lactose synthesis

The lactose in the milk produced during the 4½ hours after injection was hydrolysed, and the glucose and galactose both degraded in such a way that the specific activities of carbon atoms 1+2+3, and also 6 *in each sugar* could be determined. In each, more than 95 per cent of the activity was in carbon atoms 1, 2, 3. This experiment suggests that blood glucose does not pass into a single 3-carbon atom intermediate nor yet into a 2-carbon atom intermediate before it is synthesized into lactose by the mammary cells. If it had done so, the radioactivity in the glucose and galactose residues after hydrolysing the lactose

Grant (1936) showed, with slices of fresh active mammary gland from the guinea-pig, that lactose was actually synthesized by these slices when they were suspended in oxygenated Ringer's solution containing glucose, but not galactose or a number of other substrates. It is known that isolated surviving udders will continue to secrete lactose provided the glucose content of the perfusing blood is maintained. In such cases, the total amount of glycoprotein in the perfusing blood would be insufficient to provide the galactose needed.

From these and other experiments, it was possible to reach the tentative conclusion that lactose was produced in the mammary gland mainly from the circulating glucose of the blood plasma.

Recent experiments of former Shinfield colleagues, with their collaborators, appear to have established this conclusion very firmly. Malpress and Morrison (1950), using two independent methods for estimating lactose, have confirmed and extended the work of Grant just referred to. They have shown qualitatively by chromatography and also quantitatively that both glucose and glycogen, but not maltose nor lactate, could be synthesized into lactose by guinea-pig gland slices. Further proof has been brought still more recently (1952) by French, Popják and Malpress, using radioactive starch administered by stomach tube to lactating rabbits. Lactose was isolated from the milk produced during the next six hours. Not only was the lactose radioactive, but the specific activities of the carbon of the glucose and galactose moieties were found to be the same. It is reasonably inferred that both parts of the disaccharide molecule are formed in equal measure from the same source—glucose—and that other sources than glucose are unlikely. If, for example, the galactose combined in the blood glycoprotein had contributed substantially to the lactose molecule, this would have diluted the galactose moiety with non-radioactive material, so that its specific radioactivity would have been well below that of the glucose half.

Barry permits me to quote further unpublished work (1951) that he has done recently with Reiss in Chicago. Here glucose, labelled equally in each carbon atom with C^{14}, was injected

to provide some CH_2 groups for burning! Certainly, after injection of radioactive acetate into lactating animals, the CO_2 in the expired air contains plenty of radioactivity, but the acetate might, in the intact animal, have been oxidized in tissues of the body other than the mammary gland. Isolated mammary tissue will oxidize both acetate and glucose.

As we have seen, much of the glucose *and* the acetate, and, in fact, the amino-acids of the blood as it traverses the mammary gland, are used for synthesis. Nevertheless, it seems fairly certain that one, or perhaps more than one, of the trio—acetate, glucose, hydroxybutyric acid—provides the combustible carbon. That we cannot yet say for certain what the mammary gland normally uses for energy purposes is still a dark patch where more light is urgently required.

THE SO-CALLED 'LET-DOWN' OF MILK

One of the most dramatic happenings in the daily routine of milking a cow is the phenomenon usually called let-down. This term is a little misleading, the phenomenon is rather an active expulsion of milk from the secretory portions of the gland into the collecting sinuses close to the teat than a passive relaxation of some sphincter. Thus, in a normal cow, if the quantities of milk obtained by any constant milking procedure are measured at, say, half-minute intervals, it is found that, after a slow start in the first half-minute or so, the gland becomes turgid, the milk begins to pour out with rapidly increasing speed, usually reaching a maximum between the first and second minute, and then tailing off after, perhaps, four or five minutes in all Most, if not all, lactating mammals show this phenomenon; in the nursing human mother it is often called the 'draught'.

All the available recent evidence is consistent with the view that the let-down is brought about by the liberation, from the posterior pituitary, of a hormone which circulates in the blood and causes contraction within the mammary gland, leading to increased internal pressure, so that the action of milking or of suckling, which overcomes the resistance of the teat sphincters, will now lead to a more or less rapid discharge of milk

The evidence may be briefly summarized as follows·

would have been evenly distributed between carbon atoms 1, 2 and 3 and carbon atoms 4, 5 and 6.

From these recent experiments, there is little doubt that glucose is a normal blood precursor for mammary lactose, and that the 6-carbon-atom skeleton of the blood glucose is probably not broken down during this synthesis. However, Popják, Folley and their collaborators have shown that the goat udder can also, to an appreciable extent, transform circulating acetate into lactose, and the same ability has also been demonstrated for the perfused, isolated mammary gland of the cow.

THE SOURCE OF ENERGY FOR MAMMARY ACTIVITY

That the active mammary gland requires and uses a large amount of energy is very patent. There is a quantitatively large and continuous utilization of oxygen, as is evidenced by the rapid blood flow, together with the abstraction from that blood by the gland of up to five volumes of oxygen per cent, i.e. of the order of 20–30 per cent of the oxygen in the arterial blood, and the return to the venous blood of rather more than that volume of CO_2, giving a respiratory quotient during lactation of well above unity. Further, surviving tissue-slices from a lactating gland use more oxygen per gram per minute than most of the other main tissues of the body, with the exception of the kidney. There is a steady and quantitatively large rise in the oxygen consumption per gram of mammary tissue from the last days of pregnancy to the height of lactation, when the oxygen utilization per gram of tissue may be as much as eight times the prelactation value.

What carbon compound or compounds, then, are being oxidized in the mammary gland? It is known (see Table 1) that glucose, acetate, β-hydroxybutyric acid, neutral fat, amino-acids and glycoprotein all diminish in concentration in the plasma passing through the gland.

There have been two or three rather despairing hypotheses in the past, one that was proposed before it was known that the lower fatty acids of milk fat came from circulating acetate being that some of the long-chain fatty acids of the neutral blood fat might have been, as it were, whittled down to short-chain acids

hormone, seems to be considerably greater than would be expected from this very small content of oxytocin. It is, of course, remotely possible that a third hormone, which divides itself as an impurity unequally between the two usual fractions—the oxytocic and the pressor—of posterior-pituitary extracts may be the responsible agent. If so, it must be active in extraordinarily small amounts.

Oxytocin itself is, of course, extremely active. In the lactating sow, for example, as little as one-third of an international unit will cause measurable let-down effects. It is, perhaps, not surprising that it has not yet been directly demonstrated that the blood of an animal stimulated to let down contains more circulating oxytocin than that of the same animal *before* stimulation.

What is the contractile mechanism in the gland, acted on by the oxytocin (or other posterior-pituitary hormone), by which intra-mammary pressure is suddenly increased at let-down? It is clear that there must be what amounts to a squeezing of the mammary sponge so that the secreted milk is forced out of the alveolar spaces and small ductules into the larger ducts and collecting vessels of the gland, and one looks for evidence of contraction of unstriated muscle, of which there is some in the gland, and which, in various other organs of the body, such, for example, as the uterus, is known to respond to sizeable doses of oxytocin. But the unstriated muscle in the udder is relatively meagre in quantity and what there is of it seems to be in the wrong places, i.e. not associated with the alveoli

The missing links in the argument have been furnished by the work of Richardson over the last six or seven years (see Richardson, 1949). By improved methods of staining and other advances in technique, he has clearly demonstrated that surrounding each alveolus there is a kind of basket-work made up of the interlocking processes of certain cells whose large nuclei, lying between the cuboid milk-secreting cells of the alveoli and the basement membrane to which these secreting cells are attached, were first observed many years ago. This basket-work, myoepithelial structure is, Richardson's histological work shows, sufficient in quantity and in exactly the right place to bring about expulsion of milk both from the alveoli and the ducts if it is able to contract

1. It was shown over forty years ago that injection of posterior-pituitary extract into the circulating blood causes a flow of milk from a lactating animal whose mammary gland has been cannulated through the teat.

2. Gaines (1915) showed that such extracts would increase intra-mammary pressure. He also showed that there was appreciably more milk in the udder than was obtainable at any one milking.

3. Turner and his collaborators in the 1930's came to the conclusion that the let-down of milk was a reflex normally brought about by sensory stimulation of the teats and implemented through the central nervous system by liberation of a hormone from the posterior pituitary.

4. Petersen (1944) and his co-workers, during the early 1940's, showed that in cows in which the efferent nerves to one half of the udder had been cut, the let-down follows equally in both halves of the udder when the oxytocic hormone of the posterior pituitary is injected intravenously. They further showed that in the isolated, surviving udder, perfused by oxygenated blood from the same cow, the replacement of this blood by blood from cows just previously stimulated to let down their milk causes a copious flow of milk from the isolated udder. Blood taken from cows not so stimulated has no effect. These workers also showed that the administration of adrenalin prevents the let-down effect of the oxytocic hormone.

5. Cross and Harris have recently (1951) been able to cause let-down of milk in the lactating rabbit by electrical stimulation of the region of the pituitary stalk. The ejection of the milk followed about 25 seconds after the start of stimulation, reaching a peak after about 1½ minutes, and subsided in four to seven minutes. The response was similar to that elicited by intravenous injection of a small dose of commercial 'pituitrin'.

There can, I think, be little doubt that 'let-down' or 'draught' is caused by reflex liberation of a hormone from the posterior-pituitary gland, but whether the hormone is identical with oxytocin or not is not yet quite certain. The activity, in causing let-down, of preparations of the pressor hormone of this gland, preparations supposed to contain minimal quantities of the oxytocic

inferred that one at least of the factors leading to mastitis is too large an effective size of the teat orifice. (Heifers only were used, because it is well established that susceptibility to udder infection usually increases with age.) We are thus left with the unpleasant dilemma that if we breed and select for quick milking —a valuable adjunct to a cow from the cowman's standpoint— we are, it appears, also selecting for increased liability to udder infection.

TABLE 2. The Correlation between Machine Milking Rate and Susceptibility to Mastitis

Group	Mean Milking Rate (lb /min.)	Proportion of cows becoming infected (per cent)	Proportion of cows becoming clinical (per cent)	Proportion* contracting dry period infections (per cent)
A	2 42	10 0	5 0	15 0
B	3 56	40 0	20 0	13 3
C	4 50	46 1	19 2	23 1
D	5 55	41 7	41 7	33 3
E	6 79	66 7	44 4	44 4

* In dry period following first lactation

It must not be imagined that a normal let-down of milk, associated with as complete an evacuation of the mammary gland as possible, either by the suckling or by hand or machine, removes all, or nearly all, the milk that was originally present. It has been shown, for several lactating mammals, that after a normal milking-out, more milk can readily be obtained by giving the animal an intravenous injection of oxytocin. Thus, in a series of cows, my colleagues Dodd and Foot (1948) have shown that on the average about 20 per cent of the milk originally present in the udder of the cow remains behind after a normal milking-out. An interesting feature is that this 20 per cent contains as much as 50 per cent of the fat originally present,

on oxytocin stimulation. Similar myoepithelial cells around the ducts extend their processes in such a way that on contraction they shorten the ducts and increase the width of the lumena, thus permitting rapid movement of the milk from the squeezed alveoli. His histological evidence also appears to show that *after* let-down the processes of the myoepithelial cells have actively contracted (and not merely collapsed) with the squeezing out of the milk from the previously distended alveolus.

The case for the system of myoepithelial cells being the effective contractile mechanism which, stimulated by circulating oxytocin, forces milk into the collecting tubules and sinuses, is most convincing, though it cannot yet be regarded as completely proven.

THE REMOVAL OF MILK FROM THE GLAND

Let-down, however, is only half the process of getting milk from mother to offspring, or cow to bucket. The milk may be under pressure in the larger sinuses in the mammary gland, but before it is discharged it must get past the teat sphincter, which seems quite unaffected by the let-down process. To do this, suction, either natural or mechanical, must be applied from without.

Every cowman knows to his cost the obstinate difference between the hard-milking cow, which may take as much as ten or twelve minutes to milk, and the easy milker, which discharges her milk in three or four minutes. Recent work by colleagues at Shinfield has shown that there is little or no difference in such animals between let-down efficiency, but that the difference is in the effective size of the teat orifice. A small cannula placed in the teat canal of one quarter of the udder of a hard milker transforms that quarter at once into an easy-milking quarter.

This group of experiments has yielded some interesting observations on mastitis, which in this and other dairying countries is perhaps the chief bane of the dairy farmer. Dodd and Neave (1951) have shown, with 94 first-calf heifers, by analysing their rate of milking and the mastitis records, that there is a strong correlation between milking rate and mastitis, the faster the inherent milking rate of a cow, the more liable she is to mastitis infection (see Table 2) The milking rate is an objective measure of the patency or security of the teat sphincter, so it can be

Pavlov's findings on the establishment of such reflexes should be borne in mind by those concerned with effective human, as well as bovine, lactation. The article of Waller (1947) in volume 5 of the *British Medical Bulletin* is worth reading in this connexion.

I have been able to deal, and that rather sketchily, with two aspects only of recent research on mammary function. I have said nothing on the growth and development of the mammary

TABLE 3. Approximate Yields of Milk per Mammary Gland following the Intravenous Injection of Varying Amounts of Oxytocin in the Sow

Oxytocin (i.u.)	Yield of milk (ml.)	Oxytocin (i.u.)	Yield of milk (ml.)
1/96	0–10	3	80
1/64	0–10	5	85
1/32	10	7½	95
1/16	10	10	125
1/8	10	12½	130
1/4	20	15	155
1	40	17½	135

Approximate amount of milk obtained by a piglet at a natural suckling = 30–35 ml

gland, nor on the factors controlling the inception of lactation, nor on the so-called virgin lactation induced by oestrogens, nor on stimulation of yield by thyroxine administration, nor on grass tetany, nor on milk fever, nor on half a dozen other mammary topics on which valuable findings have been made of recent years. Labourers in the vast field of mammary physiology, biochemistry and pathology are few, but activity is developing and I personally shall be delighted if this possibly ill-assorted selection of a few recent efforts to throw light on mammary function may have tempted a potential research recruit or two to peer over the fence and at least examine the ground.

i.e. what is obtained in the residual milk, or 'super strippings' as they are sometimes called, is, in fact, a cream, containing perhaps 15 per cent of fat. When a cow is milked, the first portion of the milk obtained is poor in fat, and the fat percentage steadily rises to the end of milking, the last-drawn milk being the richest. This is also true for the other species of mammals that have been investigated.

Gross variations in fat percentage and volume per milking, which not infrequently occur in dairy cows or goats, may be associated with an insufficient or inefficient discharge of hormone by the posterior pituitary. It is well known that in certain cows (and the variation in this respect between cows is great) casual stimuli at milking time may lead to a diminished yield of milk. There is little doubt that this is due to some defect in the normal reflex, most probably at the pituitary end, where less than the normal quantity of oxytocin may be discharged. In some species, such as the pig, the amount of milk the sow gives subsequent to oxytocin injections depends on the number of units of oxytocin administered. The sow has an astonishing control over her let-down, and a normal let-down causes the liberation to her litter of only a small fraction, perhaps one-fifth or one-sixth, of the total milk in her mammary glands, as can be shown by subsequent injection of a fairly large dose of oxytocin. Table 3 gives some recent results obtained at Shinfield by Braude and Mitchell (1951).

The let-down of milk belongs to the type of physiological activity where psychological factors play a large part. Every medical practitioner is only too familiar with functional disorders, not always trivial, that may arise from causes which can be described as purely psychological. In the cow, high yields and prolonged lactation periods depend in part on the genetic make-up of the cow, but they are greatly influenced by regular milking intervals, avoidance of disturbance to the cow that may lead to anxiety or fright, or apprehension of pain, and constant gentleness of handling both of the animal and her udder. One major factor in high yields is the early establishment of a regular routine, to be departed from only at peril of lowered output. The let-down of milk is a clear case of a conditioned reflex, and

IV

On the Organization of Cortical Mechanisms

G. JEFFERSON

In what follows no attempt will be made to review the properties of the different cortical areas as disclosed either by physiological research or by clinical neurology. I should not have a great deal to add to the brief review that I made eighteen months ago (1950). In a repetition I might lay more emphasis on certain things, such as new examples of the influence of changed parameters of stimulation on motor response by Liddell and Phillips (1950) or Gellhorn and Johnson (1951) who showed that alterations in the voltage, duration or frequency of the stimulus each resulted in variations quantitatively in different muscles in the contra-lateral limb. With certain voltages, for example, contractions in some muscles would be dominant; alter the frequency and it is others which mostly reply. Always more than one muscle responded, as was proved by electromyography. Certainly stimulations of the motor cortex under modern conditions support the view that similar movements can be elicited from several different points The pre-central strip no longer seems to be a fixed mosaic from which stereotyped responses must always result. The examples of secondary motor areas reported by Adrian (1946) and Penfield (1950) expand the regions from which movements can be caused. They produce more general movements than those from areas 4 and 6. Again, something might be said of the pathfinding properties of the parietal fields in man, *on their part in orientating the body* in space (Gordon Holmes, 1919; Riddoch, 1941; Russell Brain,

REFERENCES

BALCH, C. C. (1951). Private communication.

BARRY, J. M. (1951) Private communications.

BRAUDE, R., and MITCHELL, K. G. (1951) Private communication.

COWIE, A. T., DUNCOMBE, W. G, FOLLEY, S J., FRENCH, T. H., GLASCOCK, R. F., MASSART, L., PEETERS, G. J., and POPJÁK, G. (1951). *Biochem. J.* 49, 610.

CROSS, B A., and HARRIS, G. W. (1951). *J Physiol.* 113, 35P.

DODD, F. H., and FOOT, A S (1948) *Agriculture*, 55, 238.

— and NEAVE, F K (1951) *J. Dairy Res* 18, 240.

ELSDEN, S. R., and PHILLIPSON, A. T (1948). *Annu Rev. Biochem* 17, 705.

FOLLEY, S. J., and FRENCH, T. H. (1950). *Biochem. J.* 46, 465.

FRENCH, T. H, and POPJÁK, G. (1951). *Biochem. J.* 49, *Proc. biochem Soc.* iii

— POPJÁK, G, and MALPRESS, F H. (1952). *Nature, Lond* 169, 71.

GAINES, W. L. (1915) *Amer. J. Physiol.* 38, 285

GRAHAM, W R. jnr, JONES, T S G., and KAY, H. D (1936). *Proc. Roy. Soc.* B. 120, 330

GRANT, G. A (1936) *Biochem. J* 30, 2027

MALPRESS, F H, and MORRISON, A. B (1950). *Biochem. J.* 46, 307

McCLYMONT, G L (1949). *Biochem J.* 45, *Proc. biochem Soc.* i.

PETERSEN, W E. (1944) *Physiol Rev* 24, 30.

POPJÁK, G, FRENCH, T. H., and FOLLEY, S. J (1950). *Biochem. J.* 46; *Proc. biochem Soc* xxviii.

— FRENCH, T H, and FOLLEY, S J (1951) *Biochem. J* 48, 411.

— FRENCH, T H, HUNTER, G D., and MARTIN, A. J. P. (1951). *Biochem. J.* 48, 612

REINEKE, E P, WILLIAMSON, M B, and TURNER, C W. (1941). *J biol. Chem.* 138, 83.

RICHARDSON, K C (1949) *Proc Roy Soc* B 136, 30

RITTENBERG, D, and BLOCH, K (1945). *J biol Chem* 160, 417.

SHAW, J C, and KNODT, C B (1941) *J biol Chem* 138, 287

WALLER, H (1947) *Brit med Bull* 5, 181

Instead of concentrating on this side of the subject, which could be illustrated from surgical neurology, I propose to pay attention to something different, the organization of the cortex in terms of mechanism and its relation to the brain-stem in particular. We shall see that that is not a whimsical or empirical [illegible] tion may be. I shall not pretend to give a complete explanation, since that is not yet possible, but rather to give a brief account of how matters stand today. Permit me to commence by going back a long way to trace the steps by which we have come.

CORTEX AND LOWER LEVELS

In 1897 Sherrington wrote these words: 'Even the higher psychical events cannot truly be spoken of as functions of the cortex in the sense that they are simply the outcome of molecular changes in the grey matter; they are rather to be regarded as the outcome of complex processes in which the parts of the brain below the cortex play a part no less essential than that of the cortex itself The fibres passing down from the cortex to the middle brain have probably functions by which they take part even in our psychical life, functions for which neither the words motor nor sensory are fitting.' It is to the phrase 'functions for which neither the words motor nor sensory are fitting' to which attention should be particularly directed. They readily find an echo in the minds of neurologists today, especially of those searching for the mechanisms underlying the integration of the cerebral cortex. They are used here with the special purpose of giving an indication of the lines along which this review will proceed. It is impossible to attempt to speak of cortical integration without taking into the fullest account cortico-subcortical interplay, a subject in which great gains have been made in recent years. It is interesting that Sherrington's experimental work even at that early date fortified him to modify a too wholehearted reliance on the cortex as explanatory of such things neurological as were neither spinal *nor cerebellar* Spinal cord, cerebellum, cerebral cortex—so the writ might have run after

1941; Critchley, 1950) and on the influence of the same areas in the formation of the body image, in our awareness of our own limbs and half-selves, admirably expounded by Purdon Martin (1949). Reference could also be made to the discovery by Amassian (1951) of the electrical signals of arrival in the post-central gyrus of firing by mesenteric afferents.

Aidar *et al.* (1952) working along the same lines have shown that these splanchnic afferents pass up the ordinary somatic sensory pathways—the posterior columns and medial lemniscus—or by the spino-thalamic tracts, to the contra-lateral thalamic nucleus ventralis posterior. This is what they should do if (as most of us believe) they travel over afferents of the central nervous system and do not belong to the sympathetic nervous system (as a few dissenters have contended, e.g. Foerster). Another subject about which more has emerged is the extent of cortical representation of autonomic functions. Wall and Davis (1951) have been able to distinguish three separate autonomic cortical systems, (1) the anterior half of the island of Reil which merges with the orbital cortex, and sends its efferents through the hypothalamus, (2) the sensori-motor cortex with efferents either in or accompanying the pyramidal tract, and (3) the medial temporal lobe tips which have extra-hypothalamic connexions. Stimulation in all these areas causes elevation in blood pressure and alterations in heart and respiratory rate. Mention might also have been made of the effects of stimulation of the anterior sigmoid gyri in the frontal cortex of the cat by Hoff *et al.* (1951). Repeated stimulations not only produced sharp rises in blood pressure, as much as 100 mm Hg., but, what is more interesting, ischaemia of the renal cortex. In chronic experiments where the animals were electrically shocked through the intact skull 10–15 times a day for 4–6 weeks actual fatty degeneration and necrosis of the renal tubular cells resulted. The eventual picture might closely resemble lower nephron nephrosis. All this information helps in clarifying the conception of the brain's dominant role in control of the body Now that parts of the brain can with some justification be called 'the visceral brain' to use Maclean's term (Fulton, 1951) it seems by its christening to have a more recognizable identity.

lower organisms are capable of such varied behaviour; animal semantics show that they live more different lives and have more diverse habits and abilities than their nervous systems should have permitted—or so plain anatomy seems to say. For the moment it will suffice to say that cerebral integration seemed long ago, and is beginning again to appear to be chiefly a brain-stem and diencephalic concern. Well might Sir Charles Bell have invited his readers to study Plate IV in his book *The Nervous System* (1830) with the remark, 'He who makes himself master of this plate can have no difficulty in comprehending the whole nervous system;—he holds the Key to it in his hand'. The plate is a drawing by Bell of the brain-stem from the sliced thalami above to the upper cervical segments below.

It is not such a far cry from considerations such as those to the interpretation of the cortex today. Nervous systems exist as communication systems, more particularly as information systems. The rest of our bodies exist to assist our nervous systems by providing them, firstly, with the energy required for their use and maintenance, and, secondly, by giving us the mobility that we require for finding food, for defence and for reproduction—for maintaining ourselves and our succession in a world, an environment, in which these things have to be won. But vital though those things are, our chances of success are enhanced by the opportunities our bodies afford us of making the fullest use of our nervous structures by allowing us to gather information from wide sources and to record it. What is more, we know the brain itself is interested in a long-distance control of metabolism through the 'visceral brain', hypothalamus and the pituitary apparatus. There is much more here than could have been dreamed of by those early experimenters who discovered electrically stimulable areas in the cortex. I should put it better if I said 'the investigators who first obtained visible results from cortical stimulation', for few can doubt that all areas of the cortex are stimulable, i.e. can be activated by stimulation provided that the right stimulus is used, though the results must sometimes be looked for in alterations of the heart-rate, in blood pressure, respiration, blood sugar and so forth. The island of Reil is a good example already mentioned. Or again results

the first flush of triumph of cortical stimulation. But fifty years ago it might have been easier to remember the functional values of lower levels, not only because of their brilliant and then more recent explanation by Hughlings Jackson but because, in the first three-quarters of the previous hundred years, attention had been focused precisely on these lower levels of the brain. The difficulty of our forerunners fifty years again before Sherrington had been much the same as ours. They had to fit the cap of the cerebral mantle on to the brain-stem. Their approach had been in the opposite direction from ours. Theirs was from below upwards, whilst ours is largely from above down. The reasons for the difference in these points of view were two. First, they knew that it was no good looking for specific functions in the cortex. It had been conclusively proved by Flourens, Longet and others that only the spinal cord, medulla, and corpora quadrigemina could be excited either electrically or mechanically to produce movement or convulsion, secondly, a great deal of the working out of the brain mechanisms by Bowman, Carpenter, Mayo and R. B. Todd had been by inspired guessing in the light of comparative anatomy—a good enough method within its limitations and not without its use today.

All medical students tread this path again when they study the structure of the frog, dog-fish and rabbit and observe the growing importance of the cerebral hemispheres, and the obscurations of the optic vesicles and olfactory apparatus as the most immediately striking features of the brain. We see in this history a sketch of Hughlings Jackson's conception of levels in which the acquisition of more powers by the hemispheres modified the uses, and notably the structure, of lower levels. There was no suggestion in Jackson's mind that man and the higher or more complicated creatures had a nervous system which had been built up in detachable layers. No injury, no disease could abstract everything that made man not a frog and leave him just a frog. Too much alteration had been produced in lower levels by the evolutionary process to make this a possibility—compare for instance (if an example is needed for so obvious a fact) the plan and number of cells in the *human brachial enlargement* compared with that of the frog. The astonishing thing is that

the words motor nor sensory are fitting'. Their different anatomical plans might be a sign of local use in the circuits of an electrochemical machine. When once we come to understand what such and such a cellular layout means in mechanical terms, then localization, even parcellation, may be welcomed back in quite a different guise—no longer related to intellectual or other values but to an intricate mechanism that we either have no language to describe or to which terms cannot be affixed except tentatively with symbols of interrogation after them.

As for the minute anatomy of the cerebral cortex, its cyto- and myelo-architecture, its stratification are familiar. It must be assumed that its arrangement has a functional significance in terms of mechanics though we are still feeling our way towards an understanding. We lack a blue-print of its construction largely because the profusion of cortical cells—40,000 per sq. mm (Thompson, 1899)—and, above all, of their connecting fibres, is too great for us to understand the meaning which must be there. The stratification of the cortex persuaded us of the organization of the cortex in a vertical direction, a plan that we readily accept. However it is plain from the neuropil with its vast numbers of horizontally running fibres and axonal branches that the cortex has an intrinsic organization of its own and that it ought to be regarded as a nerve-net with an integration in its own right. Yet it is quite plain to a clinician that the cortex owes its usefulness in a general sense to its communications with deeper structures, and that its function is to make use of information received, of input, if that word is preferred, and to reply to it in part immediately by efferents leading to action or by modification of activity elsewhere in the cerebral mass or by storage with the possibility of recall, memory, association and invention.

For the moment histology seems to be unable to make any further revelations and such as come must be of a different kind from the old plain study of cell morphology. Diagrams of theoretical interactions between cells have been made by Lorente de Nó [illegible]

since it could not activate enough of the 100 synaptic knobs on

must be searched for electronically until a response in the form of altered activity is found in some distant or near cortical area.

CORTICAL STRUCTURE

So far no mention has been made of histologically different areas in the brain, the charts of which held out at one time high hopes that they would supply us in good time with a new insight into local cerebral function. It is today generally agreed that cortical maps have not fulfilled the aim of Brodmann, for example, who wrote 'our goal is the development of a comparative organology of the cerebral surface, based upon anatomical criteria' (1925). No use, we think, to look for a mental quality such as a joke in the cerebral cortex unless it be directed at an effort to find one. But however reminiscent Brodmann's remark may be of the plurality of minds of the phrenologists, it was made at a time when hopes had not yet faded. Today the exertions of later workers, notably of Lashley and Clark (1946) and of Percival Bailey and von Bonin (1951) have thrown grave doubts on the validity of structural cortical maps. Objection lies not in approved homogeneity of the cortex—that is not what is found—but in variabilities such as impede the recognition of an histological slice as belonging necessarily and always to a single area. There are exceptions, of course, the Betz cell areas, the visual areas, the hippocampus, in a wider sense the granular and agranular cortex. Le Gros Clark (1952) is rightly not prepared to deny all parcellation. Whilst the most fundamental fact about the cortex is that it is so similar all over, there are quantitative differences in many regions in cell population and fibre density. The cartographers were more nearly right than present observers will admit. The fault lies not in the original observations but in those who have read since a great deal more function into the maps than the anatomists intended. Before we abandon localization altogether we should inquire whether the facts do not permit of a different interpretation. Instead of expecting to find in small focal areas a mosaic of parcellated mental or sensory functions, might it not be that areas are locally active in a particular way in facilitating or inhibiting activities elsewhere? We might use again Sherrington's words 'functions for which neither

in the cortex. He does not believe, in fact no one does, that cortical speeds of discharge depend on rhythms set by a pacemaker in the basal ganglia or elsewhere. Undercutting the cortex or removal of the thalamus (Morison and Dempsey, 1942) greatly reduces cortical activity and much work has been done to try to explain the meaning of that happening. Experiments with isolated slabs of cat's cortex have been made by Kristiansen and Courtois (1949) and Burns (1950, 1951). Provided that the blood supply of the isolate is intact, it remains active in the sense that for hours it will respond to stimuli such as strychnine, acetylcholine or electricity. There is no agreement as to whether the preparation shows spontaneous activity Kristiansen and Courtois produced tracings that showed that it did, though in attenuated form. Burns, on the other hand, could find no activity unless he left a narrow bridge connecting the isolate either to surrounding cortex or to the white matter. The isolate was alive but silent. Whichever the answer may turn out to be we can be sure that lively mechanisms exist in the cortex, thus it very likely has a quiet life of its own though in normal states it has no chance to live it so continuous is the barrage coming into it from lower levels. Intra-cortical spread of activity certainly occurs. The thalamus itself has its own rhythms but it is not itself responsible directly for every cortical wave. In fact rhythms in cortex and thalamus are often discrepant. Thalamograms have been studied in man and animals, particularly by Williams and Parsons-Smith (1950) in this country, and by Jasper (1947) and Spiegel and Wycis (1950) in North America. We may assume that 'feed-back' systems between cortex and thalamus exist. The theory of reverberating chains has become very popular, though no one has ever seen one in action. And indeed, except as momentary affairs, they are doubted by neurophysiologists This is in striking contrast to the amateurs of the cybernetic school who have postulated self-exciting chains going on and on for ever as the means by which we remember and conduct ourselves. No physiologist today believes this

The next question is whether the whole of the normal thalami firing into the cortex occurs only over classical afferent thalamo-cortical pathways (sensory, visual, auditory, from cerebellum,

each of so many nerve cells. It may be doubted whether these diagrams are as helpful as they seem at first sight. All are highly stylized and formal, at the best they remind us that nerve cells act and react on one another. The actual mass of the cortex is made up considerably more by fibres than by cells. J. Z. Young suggests that cell size may depend mostly on the number of dendritic cell connexions, i.e. on the volume of its input, though admittedly the length of the issuing axon has something to do with it. Young (1951) said 'For our present purpose the point is that there is evidence that the size reached by any nerve cell is influenced by its central and peripheral connexions and, though more speculatively, by the number of impulses that cross those junctions'. We remain for the present unable to make great use of the shapes and volumes of cortical cells, the more so since the failure of area giganto-pyramidalis to explain the motor cortex (Walshe, 1948) or the pyramidal pathway (Lassek, 1942).

ACTIVITY OF CEREBRAL CORTEX

The outstanding features of cortical activity are the fluctuations in its electrical potential in cycles per second as recorded by the E.E.G., the dominant one in the brain completely at rest but awake being the alpha rhythm at 8–13 c/sec. This activity varies little in different individuals or in the same person at long intervals. As most people know, this rhythmical activity is upset by attention, by interest in anything presented to the subject as well as by drowsiness, by sleep, and by chemical alteration in the cerebral blood flow. It is greatly altered in pathological states of depressed consciousness (parasomnia) whether experimentally produced or caused by injury or disease. In these states the fast rhythm becomes slower, the fluctuations of potential more irregular and the voltage higher The effect is diffuse and symmetrical over both hemispheres. It cannot be produced by local cortical injury but is easily caused by stimulations or injuries to the brain-stem or to some parts of the thalamic nuclear complex. The question is whether the cortex has an intrinsic rhythm of its own or whether it owes it to projection fibres from lower levels. Some, such as Bremer, believe that auto-rhythmicity is a property of highly organized collections of nerve cells such as exist

red nucleus, etc.) or whether there is something else for which no name has existed. The point is one of considerable clinical interest. Its elucidation has been helped to some extent by the discovery by Bremer (1935, 1936, 1949) of two new laboratory preparations—'le cerveau isolé' and 'l'encephale isolé'. The former consists in a brain in which the brain-stem has been severed with a blunt spatula at tentorial level. The other is a brain isolated by dividing the spinal cord at the first or second cervical segment. These will be recognized as the same preparations as those which Sherrington used for the study of decerebrate rigidity and spinal reflexes. Bremer started with the hypothesis that sleep in mammals, whether natural or pathological, depended on more or less complete de-afferentation. This view would be quite consistent with Kleitman's (1939) well-known researches on sleep. Bremer thought that subcortical structures produced 'tone' in the cortex, a readiness to act. Bremer's hypothesis turned out to be correct for the large slow waves with spindle bursts of activity characteristic of deep sleep were obtained from the cortex of the cerveau isolé. In brains with sections at C.1 the head remained awake but peaceful with low-voltage fast activity in the cortex. We shall return to the explanations proposed by Magoun and his collaborators in a moment. Meanwhile Jasper in Montreal had pursued the problem of the influence of the thalamus on the cortex. His experiments led him to propound the existence of a thalamic reticular system with a diffuse cortical projection that alerted the cortex and in fact was a great factor in integrating the brain. The projection did not travel over the classical thalamo-cortical pathways which might or might not be activated concurrently. The evidence that Jasper produced is too extensive for detailed discussion. But one set of experiments is worthy of further comment, that which has to do with the possible nature of petit mal.

Stimulations of various parts of the thalamus and hypothalamus were undertaken to investigate the hypothesis that the diencephalon might be the site of origin, or at least of major activation, of the petit mal attack. A strong suggestion that had emerged from previous work by Jasper with Droogleever-Fortuyn (1947) was that the pace-maker for the characteristic

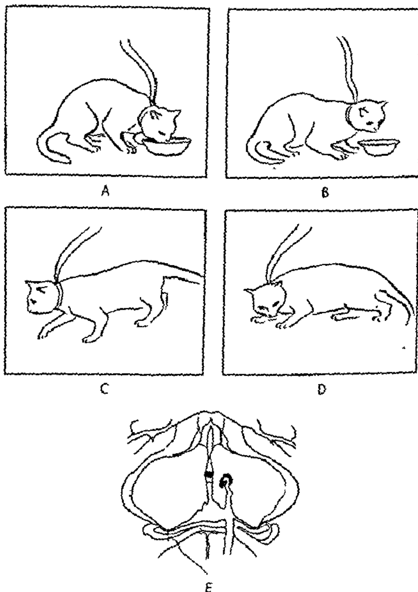


FIG. 1. A to D The 'arrest' phenomenon, sudden medial thalamic stimulation of cat drinking and walking E shows site of stimulus (By courtesy of J. Hunter and H. Jasper)

linked most closely with the anterior temporal lobes. The conclusion is not to be drawn that actual disease will be found in patients with petit mal, in fact it is a well-known clinical fact that it is a rare, but not impossible, form of epilepsy in local brain disease or injury. But it seems very probable, as Hunter and Jasper claim, that the thalamic reticular formation is the most important link in the chain, in activating both hemispheres

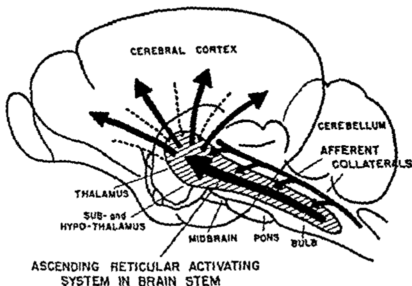


Fig. 2 Diagram of extent of reticular formation—the 'arousal' area, connexions, spinal afferents and diffuse projection to cortex (By courtesy of H. W. Magoun.)

instantly and equally and causing the 'arrest' or petit mal picture—one in which muscular tone is often not affected. It has not been found possible to turn this observation to any use in the treatment of petit mal Spiegel, Wycis and Reyes (1951) have reported on six patients in whom small electrolytic lesions were made in the lamina medullaris interna of the thalami and in none were the attacks abolished They remarked that although it is quite true that diencephalic discharges can always be recognized in the thalamograms of petit mal patients, they were unable to find a sharply circumscribed focus common to all sub-

cortical discharge seen in attacks of petit mal lay in the diencephalon. Hunter and Jasper buried electrodes in various thalamic and subthalamic sites. The leads were brought out subcutaneously allowing of repeated observations on the sleeping, waking or walking cat many days after the operation. They found that if the stimulation were made within the intralaminar nuclei of the thalamus, especially near the massa intermedia, an 'arrest' reaction was produced. By this they meant that the cat stopped in its walk, standing as if frozen in its tracks in a fixed or crouched position 'with a staring' or rather 'stunned' facial expression, to continue walking as soon as the stimulus was withdrawn. Occasionally the 'arrest' would outlast the stimulus a few seconds or longer. Slight twitchings of the lids and mouth were sometimes seen. There was no response during the 'arrest' stage to visual or auditory stimuli. Changes in the parameters of stimulation might add turning and maintenance of abnormal fixed postures to the 'arrest'. In some animals more intense stimulation precipitated generalized tonic-clonic convulsions after the initial 'arrest', confirming Hughlings Jackson's belief that both major and minor epilepsy were but different degrees of the same thing. During the arrest the E.E.G. showed the typical slow 3 per sec. spike and wave patterns of petit mal, a type that has not yet been found by stimulation of the cortex, whatever the parameters used.

Morison and Dempsey (1942) had shown that stimulation of the thalamus controlled the spontaneous electrical activity of the cortex in a diffuse and wide way. Hunter and Jasper concluded that a diffuse projection from the reticular and intralaminar nuclei (the 'thalamic reticular system') might be deeply involved in integration of the higher level cerebral activity such as the initiation of patterns of movement and that stimulation there might block such movements. In contrast to 'arrest', rage reactions and attack were seen on stimulation of the anterior hypothalamus, whilst sleep might follow stimulation of the inferior parts of the massa intermedia. Stimulations deep to the thalamus and anteriorly never produced 'arrest' but often the impulsive searching about, as if for something forgotten, characteristic of psycho-motor epilepsy, a state that seems now to be

But it still remained to be demonstrated that the afferent paths from the cord and elsewhere made connexions with the cells of the reticular formation, a supposition for which there was already some anatomical verification. This confirmation has now come and the hypothesis of cortical 'arousal' by the reticular formation looks like an established fact. It will be recognized that this view, if it is accepted, modifies considerably the theories of Kleitman and Bremer that arousal was due to a plain

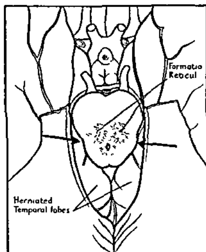


FIG. 3. Tentorial herniation, compression of mid-brain by temporal lobes —ischaemia of formatio reticularis (schematic).

sensory barrage. In one sense it is still true but there is a difference of interpretation of the same general facts. The modification lays the main emphasis on an intermediary apparatus The sensory barrage activates the reticular formation as well as passing into the thalamus and up to the cortex locally; there is both general cortical arousal and local sensory cortical recognition. Lesions of the afferent pathways from the spinal cord sparing the reticular formation did not inhibit 'arousal' E.E G. patterns on stimulation of the reticular formation. Lindsley, Bowden and Magoun (1949) found that acute upper brain-stem lesions in such medial positions as to interfere with the reticular formation had an effect the reverse of arousal. They abolished the wakeful

jects. Large lesions would be needed if all possible starting points were to be put out of action. Turner McLardy's anatomical critique (1951) of the minute anatomy of the regions stimulated by Hunter and Jasper suggests that instead of a diffuse non-specific recruiting system depending on stimulation of intra- [illegible]

McLardy's thesis that the centromedian thalamic complex fires into the other thalamic nuclei and that the final projection comes from them. This would clarify the mechanism of integrating projection. It does not, of course, deny that it occurs. Jasper's theory that there is another system of relays to the cortex is the more attractive and has the support of Magoun.

Additional information emerges from the work on the brain-stem reticular formation by Magoun and his collaborators (Moruzzi and Magoun, 1949; Lindsley, Schreiner, Knowles and Magoun, 1949; Starzl, Taylor and Magoun, 1951). These researches have much bearing on the problems of cerebral integration as a whole so far as the two great rhythms of waking and sleeping are concerned—the first with its fast low-voltage activity, the latter with its slower and larger waves. The earlier papers announced the fact that stimulation of the brain-stem reticular formation produced arousal of the whole cortex. It led to the same break-up of synchronized cortical activity with substitution of low-voltage fast activity as occurs in an animal waking from sleep or again when attention is aroused by visual or other stimuli. There appeared to be a system of ascending relays in the reticular formation which activated the cortex generally and diffusely. The pathway was through the ventromedial thalamic nuclei. The extent of the reticular formation is shown in the diagram. Magoun holds that the desynchronization of the cortex by visual, auditory, or any stimulus arousing attention is not due to a barrage of impulses arriving in the appropriate local cortical area and being spread from there inter-cortically. He believes that the barrage also excites the reticular formation which there arouses the cortex as a whole.

But it still remained to be demonstrated that the afferent paths from the cord and elsewhere made connexions with the cells of the reticular formation, a supposition for which there was already some anatomical verification. This confirmation has now come and the hypothesis of cortical 'arousal' by the reticular formation looks like an established fact. It will be recognized that this view, if it is accepted, modifies considerably the theories of Kleitman and Bremer that arousal was due to a plain

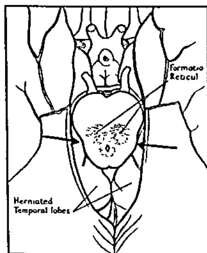


FIG 3. Tentorial herniation, compression of mid-brain by temporal lobes —ischaemia of formatio reticularis (schematic).

sensory barrage. In one sense it is still true but there is a difference of interpretation of the same general facts. The modification lays the main emphasis on an intermediary apparatus. The sensory barrage activates the reticular formation as well as passing into the thalamus and up to the cortex locally, there is both general cortical arousal and local sensory cortical recognition. Lesions of the afferent pathways from the spinal cord sparing the reticular formation did not inhibit 'arousal' E.E G. patterns on stimulation of the reticular formation Lindsley, Bowden and Magoun (1949) found that acute upper brain-stem lesions in such medial positions as to interfere with the reticular formation had an effect the reverse of arousal. They abolished the wakeful

pattern and substituted slow E.E.G. waves and spindle bursts like those of sleep or barbiturate anaesthesia.

That some such structure and organization existed had been most strongly suggested by the experiments of Ingram, Barris and Ranson (1936), and Ranson (1939), who produced long-lasting somnolence by median basal lesions that did not sever classical tracts but might well have injured such a system as the reticular. Direct damage to the reticular formation at different levels in Magoun's laboratory has resulted in somnolence. Damage led to somnolence, excitation to arousal.

That physiological mechanisms like these with a similar, if not exactly identical, anatomical substratum exist in man is highly probable. It has been evident for some years that the critical areas for arousal or wakefulness, drowsiness or coma lie not primarily in the cortex but in the subthalamic region and brain-stem. To have an anatomically intact cerebral cortex is not sufficient to ensure wakefulness, the maintenance of which seems to demand the physiological integrity of much deeper structures. That this is so is evident from the effects on consciousness of the brain-stem caused by tentorial herniation in cerebral lesions and by the seeming paradox of unconsciousness caused by acute posterior fossa lesions, especially haemorrhages (Jefferson and Johnson, 1950). Physiological blocking at this level would explain many of the paradoxes of concussion (Jefferson, 1943).

CONCLUSION

It has been said of the cortical capillaries that if a red-cell had motor power of its own it could travel from one end of the brain to the other. How much more true must this be of nervous impulses. But although a high value must be attached to intra-cortical interplays, we cannot look at them abstracted from a continuous barrage from subcortical structures Without this, cortex has no meaning. Amongst the more important of the newer discoveries of such integrating factors are those relating to the reticular formation. The older anatomists like Bell, and the classical neurologists like Hughlings Jackson and Todd and Marshall Hall, would have welcomed and understood them.

REFERENCES

ADRIAN, E. D., and MATTHEWS, B H. C. (1934). The Berger rhythm: potential changes from the occipital lobes in man. *Brain*, **57**, 355.

— (1936). The spread of activity in the cerebral cortex. *J. Physiol.* **88**, 127.

— (1947). *The Physical Background of Perception.* Oxford.

AIDAR, O., GEOHEGAN, W. A., and UNGEWITTER, L. H. (1952). Splanchnic afferent pathways in the cerebral nervous system *J. Neurophysiol.* **15**, 131.

AMASSIAN, W. (1951). Cortical representation of visceral afferents *J. Neurophysiol.* **14**, 433.

BAILEY, P., and BONIN, G. VON (1951). *The Isocortex of Man.* Chicago Univ. Press.

Paris, **118**, 1235

— (1936). Activite électrique du cortex cérébral dans les états de sommeil et de veille chez le chat. *C. R Soc. Biol., Paris*, **122**, 464.

E

E

111, 50.

CLARK, W. E. LE GROS (1952) A note on cortical cyto-architectonics. *Brain*, **75**, 96

CRITCHLEY, M (1950) The body image in neurology. *Lancet*, **1**, 335

ECCLES, J. C. (1951). Hypothesis relating to the brain-mind problem *Nature*, **168**, 53.

GELLHORN, E, and JOHNSON, D A (1950). The validity of the concept of multiplicity of representation in the motor cortex *Brain*, **73**, 267.

HOFF, E C., KELL, J. F, HASTINGS, N D. M., and GRAY, E. H (1951). Vasomotor, cellular and functional changes produced in kidney by brain stimulation. *J Neurophysiol* **14**, 317.

experimental study. *Arch. Neurol. Psychiat, Lond* **35**, 1175.

JASPER, H. H, and DROOGLEEVER-FORTUYN, J. (1947). Experimental studies on the functional anatomy of petit mal epilepsy *Res Publ. Ass nerv ment. Dis* **26**, 272

JEFFERSON, G. (1943) The nature of concussion. *Brit med J* **1**, 6.

— and JOHNSON, R. T. (1950) The cause of unconsciousness in posterior fossa compressions *Folia psychiatr. Neerl.* **53**, 2.

pattern and substituted slow E.E.G. waves and spindle bursts like those of sleep or barbiturate anaesthesia.

That some such structure and organization existed had been most strongly suggested by the experiments of Ingram, Barris and Ranson (1936), and Ranson (1939), who produced long-lasting somnolence by median basal lesions that did not sever classical tracts but might well have injured such a system as the reticular. Direct damage to the reticular formation at different levels in Magoun's laboratory has resulted in somnolence. Damage led to somnolence, excitation to arousal.

That physiological mechanisms like these with a similar, if not exactly identical, anatomical substratum exist in man is highly probable. It has been evident for some years that the critical areas for arousal or wakefulness, drowsiness or coma lie not primarily in the cortex but in the subthalamic region and brain-stem. To have an anatomically intact cerebral cortex is not sufficient to ensure wakefulness, the maintenance of which seems to demand the physiological integrity of much deeper structures. That this is so is evident from the effects on consciousness of the brain-stem caused by tentorial herniation in cerebral lesions and by the seeming paradox of unconsciousness caused by acute posterior fossa lesions, especially haemorrhages (Jefferson and Johnson, 1950). Physiological blocking at this level would explain many of the paradoxes of concussion (Jefferson, 1943).

CONCLUSION

It has been said of the cortical capillaries that if a red-cell had motor power of its own it could travel from one end of the brain to the other. How much more true must this be of nervous impulses. But although a high value must be attached to intracortical interplays, we cannot look at them abstracted from a continuous barrage from subcortical structures. Without this, cortex has no meaning. Amongst the more important of the newer discoveries of such integrating factors are those relating to the reticular formation. The older anatomists like Bell, and the classical neurologists like Hughlings Jackson and Todd and Marshall Hall, would have welcomed and understood them.

V

The Effects of Haemorrhage on the Cardiovascular System in Man

O. G. EDHOLM

HAEMORRHAGE has been widely studied both in animals and man and the effects clearly described by many workers for many years. Only a brief account will be given of the researches performed prior to 1939, as the subject has been reviewed so frequently.

The greater part of the earlier work was carried out using animals, and investigations in man were largely confined to clinical studies. During and since World War II the effects of experimental haemorrhage in man has been followed in considerable detail: man has replaced the dog as an experimental animal. One result of this work has been to emphasize the considerable differences between man and other animals. The account given here will to a large extent be confined to the results of work in man carried out in the last twelve years.

A proper understanding of the effects of haemorrhage demands a knowledge of the physiology of the circulation and of haemodynamics. Certain aspects of the circulation are of particular importance, especially the distribution of blood in the body at rest. The greater part of the blood volume is found in the veins; Fig. 1 shows the percentage distribution of blood in the veins, arteries and capillaries, including arterioles with arteries and venules with veins. If all the small vessels are included together, the proportions are, arteries 15 per cent, arterioles, capillaries and venules 19 per cent, veins 66 per cent These figures are derived from Green (1950). The veins hold nearly twice as much blood as the rest of the vascular system put together.

KLEITMAN, N. (1939). *Sleep and Wakefulness*. Chicago Univ. Press.
KRISTIANSEN, K , and COURTOIS, G. (1949). Rhythmic electrical activity from isolated cerebral cortex. *Electroenceph. clin Neurophysiol* **1**, 265
LASHLEY, K. S , and CLARK, G. (1946). The cytology of the cerebral cortex of Ateles. *J. comp Neurol* **85**, 223.
LASSEK, A. M. (1942). The pyramidal tract. *J. nerv. ment Dis.* **95**, 721.
LIDDELL, E. G. T., and PHILLIPS, C. G (1950). Thresholds of cortical representation *Brain*, **73**, 125
LINDSLEY, D. B., SCHREINER, L. H., KNOWLES, W. B , and MAGOUN, H. W. (1950). Behavioural and E.E G. changes following chronic brain stem lesions in the cat. *Electroenceph clin Neurophysiol.* **2**, 483.
LORENTE DE NÓ, R. (1934). *J. Psychol Neurol* **46**, 113.
MARTIN, P. (1949). Consciousness and its disturbances considered from the neurological aspect *Lancet*, **1**, 48.
MCCULLOCH, W. S., and PITTS, W. (1944). A logical calculus of the ideas immanent in nervous activity. *Bull. math. Biophys.* **5**, 115
MCLARDY, T. (1951) Diffuse thalamic projection to cortex; an anatomical critique. *Electroenceph. clin. Neurophysiol* **3**, 183.
MORUZZI, G., and MAGOUN, H. W. (1949). Brain stem reticular formation and activation of the E E G. *Electroenceph. clin. Neurophysiol.* **1**, 455.
MORISON, R. S , and DEMPSEY, E. W (1942). A study of thalamo-cortical relations. *Amer. J. Physiol.* **135**, 281.
PENFIELD, W. (1950) The supplementary motor area in the cerebral cortex of man. *Arch. f. Psych. u Zeitsch Neurol.* **185**, 670
RANSON, S. W. (1939) Somnolence caused by hypothalamic lesions in the monkey. *Arch Neurol Psychiatr., Lond.* **41**, 1.
RIDDOCH, G (1941). Phantom limbs and body shape. *Brain*, **64**, 197
SHERRINGTON, C. S (1897). 'The Nervous System' in Foster, *Textbook of Physiology*
SPIEGEL, E A , and WYCIS, H T. (1950). Thalamic recordings in man. *Electroenceph clin. Neurophysiol* **2**, 23
— WYCIS, H T., and REYES, U (1951) Diencephalic mechanisms in petit

cephalon *J. Neurophysiol* **14**, 461.
— TAYLOR, C W , and MAGOUN, H. W. (1951) Collateral afferent excitation of reticular formation of brain stem. *J. Neurophysiol.* **14**, 479
THOMPSON, H B (1899) *J comp Neurol.* **9**, 113
WALSHE, F M. R (1948) *Critical Studies in Neurology* Edinburgh.
WALL, P. D , and DAVIS, G D. (1951). Three cerebral cortical systems affecting autonomic function *J. Neurophysiol.* **14**, 507.
WILLIAMS, D., and PARSONS-SMITH, G (1950) Cortical rhythms not seen in the E.E.G *Brain*, **73**, 191.
YOUNG, J. Z. (1951). Growth and plasticity. *Proc Roy. Soc. B.* **139**, 18.

There are three main areas, pulmonary, splanchnic, muscle and skin, each holding approximately 30 per cent of the total volume. Skin probably accounts for some 7 per cent. The blood volume of the liver in animals, estimated by a variety of different

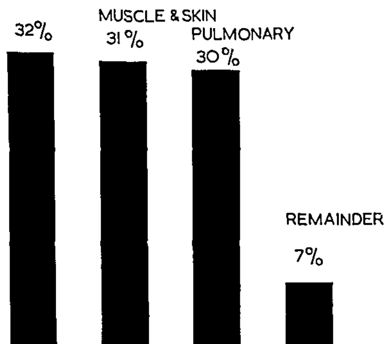


Fig. 2 The volume of blood contained in the various organs Splanchnic includes liver, intestines and kidneys These figures are derived from many sources (see text).

methods, accounts for 25–30 per cent of the total blood volume. If a similar relationship holds for man it follows that the liver blood volume would be of the order of 1500 ml. As the post mortem weight of the liver averages 1500 grammes, this implies a very high proportion of blood to cells. It would be most useful if more data could be obtained for man. The figures given in the diagram are the best obtainable at present, but it is quite

The distribution of blood in the various organs and tissues of the body is the next problem to consider. It is not quite so easy to give figures with confidence in answer to this question and much more work is required. There does not appear to be any

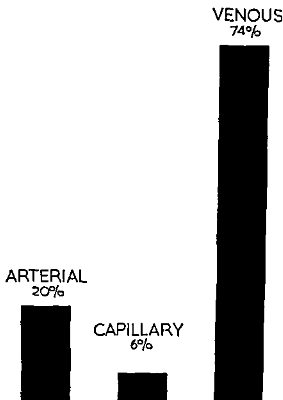


FIG 1. The volume of blood contained in the arteries, capillaries and veins, expressed as a percentage of the total blood volume. Arterioles included with arteries and venules with veins.

accurate information concerning the blood volume of the liver and intestines, i.e. the splanchnic area, in man. The figures shown in Fig 2 must be regarded as tentative only. They are derived from various sources, e.g. Ranke (1876), Silverskiold (1938), Bazett (1950), etc, and recent studies on the distribution of radioactive isotopes in the different tissues of the dog (Gibson *et al.*, 1946).

A better comparison of the vascularity of the different regions of the body can be made by expressing the blood flow in terms of unit volume of tissue. In Fig. 4 the results of such a calculation are shown. The blood flow through muscle is considerable when the total flow is considered as in Fig. 3, but muscle forms such a large mass of the whole body that, when flow is expressed per unit volume, it will be seen that muscle flow is very small indeed

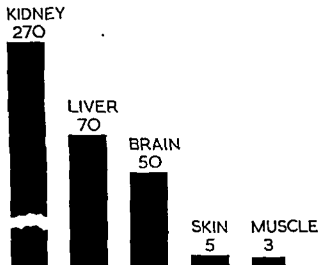


Fig. 4. Blood flow expressed as ml/100gr. tissue/minute. Kidney blood flow is so large that it cannot be conveniently fitted into the graph. It is therefore shown as broken column.

compared with the kidney or liver. The implication is that the resistance to blood flow in muscle blood vessels is high compared with the renal vessels which have a blood flow 150 times as great as in muscle (per unit volume of tissue). Resistance to flow is also low in the hepatic and cerebral vessels.

EFFECTS OF HAEMORRHAGE IN MAN

The effects of haemorrhage can now be considered in detail bearing in mind the facts outlined above. When blood is lost the volume of circulating blood diminishes Compensation for a diminished blood volume can be achieved in a variety of ways.

possible that in man the liver blood volume is considerably smaller than in animals.

Information on the volume of blood flowing through the various regions of the body is required in order to complete the scheme of the circulation. There is very considerable information available to draw up a balance sheet of the cardiac output.

SPLANCHNIC & KIDNEY
485 %
MUSCLE & SKIN
225 %
BRAIN
135 %
REST
155 %

Fig 3 The distribution of the cardiac output to the different regions of the body, expressed as a percentage of the output at rest.

Fig 3 gives the results in a highly simplified form. Complete details are given by Green (1950) showing the different values obtained by various workers. The diagram shown here is only intended to give a very broad picture. Nearly half the cardiac output goes to the splanchnic region, if the renal blood flow is included; a quarter of the output goes to muscle and skin, and the remaining quarter is divided between the brain and the rest of the body. Since the pulmonary circulation is separate from the systemic circulation the lungs are naturally excluded from this list.

been concluded that rapid haemodilution may take place in man, but Wallace and Sharpey-Schafer (1944) clearly demonstrated in man that restoration of a 20 per cent blood loss was not achieved on the average until twenty-four hours had elapsed.

Grant and Reeve (1951) in their studies on the effects of injury showed that haemodilution was a relatively slow process

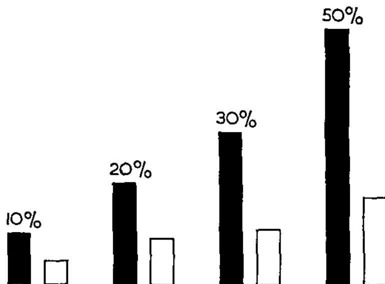


FIG. 5. The rate of haemodilution twelve hours after haemorrhage of varying degree Data recalculated from Grant and Reeve (1951), Wallace and Sharpey-Schafer (1941)

in man. The volume of fluid absorbed into the circulation in unit time was related to the volume of blood lost This relationship is shown in Fig. 5 The black columns indicate the blood volume lost by haemorrhage, and the volume of fluid restored to the circulation in twelve hours is given in the white columns. Wallace and Sharpey-Schafer's results are shown in the 20 per cent blood loss column. Their figures are comparable with those of Grant and Reeve, although in the one case blood loss was accompanied by trauma and in the other an uncomplicated venesection was performed. This supports the general thesis put

The blood volume may be restored either completely or partly by the expulsion of blood from depôts or reservoirs. A more gradual restoration can be achieved by withdrawal of fluid from the extracellular fluid into the blood.

A diminished blood volume implies a fall in cardiac output, which will result in a fall of blood pressure unless the peripheral resistance is raised by vasoconstriction. Heart rate will increase, if blood pressure falls or tends to fall, probably via a carotid sinus reflex. All of these mechanisms have been demonstrated as part of the effects of haemorrhage either in animals or man. They require to be examined in detail as regards their relative importance in man.

The work of Barcroft (1934) and his associates demonstrated very clearly the importance of the spleen as a blood reservoir in animals, including the cat and the dog. In these animals the spleen contracts after haemorrhage and the blood contained in it is expelled into the general circulation. A considerable volume of blood amounting to some 15 per cent or more of the total blood volume is found in the spleen, so small haemorrhages may be completely compensated.

There is now a substantial body of evidence to show that in man not only is the spleen of little importance as a blood reservoir, but there are no other true reservoirs (Nylin, 1947). It is probable that the large venous blood volume acts as a form of reservoir, and that venoconstriction reduces the volume of blood contained within the veins. There is a fall of pressure in the right auricle which is proportional to the volume of the haemorrhage (Warren *et al*, 1945, McMichael, Sharpey-Schafer, 1944). Silverskiold (1946) has demonstrated venoconstriction of the vena cava in animals after haemorrhage. It is necessary to emphasize the lack of importance of the spleen. It is the first main modification of the classical account so far as man is concerned.

The second mechanism for restoring the blood volume is by haemodilution. The rate of haemodilution varies considerably in different animals (Courtice and Gunton, 1949), in the rabbit 20 per cent volume reduction will be completely restored within two hours. In the dog it may take four hours It has therefore

lying down the heart rate increase is frequently trivial. There may be a greater increase during the haemorrhage itself, but as soon as venesection stops the heart rate returns towards the control level. It should be made clear that these observations regarding heart rate only apply in the absence of any fainting reaction, and with the subject lying down.

Haemorrhages larger than 20 per cent are frequently but not invariably associated with an increased heart rate. Grant and Reeve conclude that a pulse rate of 100 or more is a good indication that the blood volume is below 80 per cent of normal, but that a pulse rate of less than 100 does not necessarily imply that the blood volume is 80 per cent or more of normal.

Blood pressure changes are in many cases slight with small haemorrhages, excluding those who faint. Systolic pressure may fall 9 mm Hg with unchanged diastolic after a venesection of 500 ml (Shenkin *et al.*). In a series of volunteers who were bled an average of one litre the mean blood pressure only fell by 6 mm Hg (Edholm, 1949). These small changes in the presence of a reduced cardiac output demonstrate a marked increase in peripheral resistance.

The effects of severe haemorrhages associated with war wounds have been examined by Cleghorn and Chute (1949), Grant and Reeve (1951), Emerson and Ebert (1945), amongst others The majority of cases were seen some hours after injury. A systolic blood pressure of 100 mm Hg or higher indicated a blood volume of 70 per cent or more. If the systolic pressure was over 140 mm Hg, the likelihood of the blood volume being at least 80 per cent of normal was about 5 to 1. These figures emphasize that blood pressure is sustained in most cases even with severe haemorrhage, although a greater fall of pressure would usually be observed immediately after haemorrhage.

There must therefore be considerable vasoconstriction with changes in blood flow according to the site of the constriction. A characteristic description of a patient who has had a severe haemorrhage would emphasize pallor, cold extremities and sweating. The pallor and the cold extremities have been interpreted to indicate a great reduction in blood flow of the skin and, by implication, of the muscle blood vessels also. Emerson

forward by Grant and Reeve that injury does not interfere with the haemodilution, except in cases of abdominal or chest wounds.

It is clear that haemodilution occurs in man, but is relatively a slow process; less than half the volume of blood lost will be replaced in twelve hours. This point needs emphasizing as the rate of haemodilution in most animals is much greater, and may be described as the second modification of the classical account of haemorrhage. The immediate effect of haemorrhage is a reduction in blood volume, and in man this blood volume is not replaced from blood depôts and only very slowly by haemodilution.

The cardiovascular changes may include a fall in cardiac output, an increased heart rate, peripheral vasoconstriction and an alteration in the distribution of blood flow. The cardiac output changes are to a certain extent controversial. McMichael and Sharpey-Schafer (1944) have demonstrated a relationship between right auricular pressure and cardiac output. Small venesections of the order of half a litre lower the right auricular pressure, and cardiac output falls by some 10–15 per cent. Warren *et al.* (1945) confirmed the fall in right auricular pressure but were not able to detect a significant change in cardiac output with this order of haemorrhage. Both these groups of workers used the cardiac catheter technique. Shenkin *et al.* (1944) employed the ballisto-cardiograph and found that there was a slight and probably insignificant decline in cardiac output with 500 ml haemorrhages, but with larger venesections of one litre the cardiac output diminished significantly. It is clear that a reduction of 20 per cent blood volume or more by haemorrhage lowers cardiac output, although the effects of smaller losses are not certain. Larger haemorrhages may depress cardiac output to much lower levels (Cournand *et al.*, 1943). Animal experiments provide much data on this point (Wiggers and Ingraham, 1946).

An increased heart rate has for many years been accepted as one indication of haemorrhage. However with small vene-

[illegible] change in

[illegible] marked

[illegible] bject is

The evidence regarding skin blood flow is not so precise. *However, hand blood flow, which may be taken to represent* principally cutaneous circulation, does not diminish during venesection of 500–1000 ml. Observations were not continued after venesection ceased, and it is possible that vasoconstriction would then have been evident. In some cases hand blood flow actually increased during venesection which may have been due to emotional causes.

There is also experimental evidence of the effect of haemorrhage on renal and hepatic flow. De Wardener and McSwiney (1951) have examined the changes in renal blood flow during haemorrhage and fainting in man. Their records show that during haemorrhage and before blood pressure fell as a result of haemorrhage there was a considerable reduction in renal blood flow. Van Slyke (1948) has shown that haemorrhage in the dog is accompanied by an acute fall in blood flow through the kidney.

Bearn *et al.* (1951) made observations on the splanchnic blood flow in man during haemorrhage and fainting using the bromsulphalein technique. Their records also show that prior to fainting, i e. during haemorrhage, there was a consistent reduction in blood flow through the liver. The evidence that is available at present indicates that as a result of moderate haemorrhage of the order of 20 per cent of the total blood volume, the muscle blood vessels are not constricted, the skin vessels possibly are, but that there is a considerable increase in the resistance of the renal and splanchnic vessels. As a result of more severe haemorrhages, of 30 per cent, 40 per cent or more of the total blood volume, it is probable that there will be a reduction in the muscle-skin blood flow as well. Direct measurements in patients with this degree of haemorrhage need to be made. In Fig 6 is shown the possible changes in blood flow to the different regions as a result of a haemorrhage of the order of 20 per cent.

Some of the points mentioned already can be recapitulated in order to present a more complete account. Experimental studies with uncomplicated haemorrhage in healthy subjects demonstrate that blood loss up to one litre is tolerated very well in those subjects who do not faint. Blood pressure may be normal,

and Ebert (1945) who examined civilian casualties state that cold extremities and pallor were almost invariably observed in their cases. Grant and Reeve found that patients with these signs also had other signs of vasoconstriction, including narrowing of the radial artery and constriction of the superficial veins. Such signs were in general observed in those with severe blood loss. It is not therefore surprising that it has generally been concluded that vasoconstriction of the muscle and skin vasculature is one of the important compensatory mechanisms in haemorrhage (Richards, 1948). Teleologically, such a mechanism has much to recommend it. Not only is there a considerable blood volume in the skin and muscles but there is a considerable blood flow as well (Figs. 2, 3). These are tissues with low metabolic demands at rest, and it might be considered that blood could well be spared from such regions in order to conserve the blood supply of more essential organs such as the liver, brain, etc. However, if Fig. 4 be examined again it will be clear that the blood flow per unit mass is very small for muscle and only slightly greater in the skin, so the resistance to blood flow in these regions is very high, in comparison with the vascular resistance of the splanchnic area. On theoretical grounds it might be predicted that resistance would be increased as a result of peripheral vasoconstriction, not in the muscle and skin vessels, but in the renal and splanchnic vessels. The muscle and skin vessels are in fact already highly constricted. If the blood flow through the legs and one arm is cut off by arterial occlusion cuffs, there is very little change in blood pressure. Increasing the resistance in an area of already high resistance will have little effect on the overall peripheral resistance.

Experimental work supports this conclusion. A group of volunteers were bled, on the average one litre of blood, and the forearm blood flow was measured before and during haemorrhage and for 3½ hours after the end of venesection (Edholm). Averaging all the results showed that there was virtually no reduction in forearm blood flow as a result of a blood loss of approximately 20 per cent of the total blood volume.

Forearm blood flow can be taken to represent essentially muscle blood flow.

There are many ways of eliciting the vaso-vagal syndrome apart from haemorrhage including anoxia, passive tilting into the upright position and emotional stimuli. Fainting is not therefore a specific response to haemorrhage, and it has not yet been possible to determine the afferent impulses involved.

However, it has been clearly demonstrated that the number of subjects who faint as a result of venesection varies as the volume of blood withdrawn increases. A venesection of 500 ml [illegible] litre will almost invariably do so on standing (Shenkin *et al.*).

The cardiovascular changes in fainting have been investi-g[illegible] resistance, i.e. a marked peripheral vasodilatation. This has been shown most clearly in the muscle blood vessels of the fore-arm, where blood flow may double although blood pressure can have fallen to a mean pressure of 40–50 mm Hg. There is also evidence of a splanchnic vasodilatation, including the renal vessels (de Wardener and McSwiney, 1951; Bearn *et al.*, 1951). Cardiac output measurements show that fainting can take place with little or no fall in cardiac output, although there is frequently a decrease (Shenkin *et al.*, 1944, Barcroft *et al*, 1944; Warren *et al.*, 1945).

The onset of fainting is sudden and the recovery is usually rapid. The duration of fainting, as measured by the period of very low blood pressure, is brief, lasting from three to ten minutes, usually less than five minutes. However there are certain features which persist, including pallor and a slow pulse. Facial pallor may last up to one hour and the pulse rate seldom returns to the pre-faint level until many hours have elapsed.

It will be evident that this apparently transient episode may complicate the symptoms of haemorrhage in a subject who is seen for the first time after haemorrhage plus fainting. Indeed the relative lack of cardiac acceleration in battle casualties with severe blood loss may be partly explained by a vaso-vagal attack

pulse rate scarcely changed and facial colour good, always provided that the subject is in the recumbent position.

Examination of the records given by Grant and Reeve, Cleghorn and Chute, Emerson and Ebert show occasional cases of

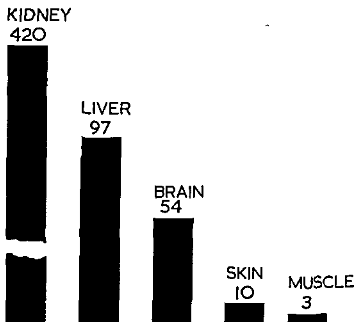


FIG. 6. Redistribution of blood flow as a result of moderate haemorrhage.

blood loss of 30–40 per cent in whom the blood pressure is substantially normal with a pulse rate of 70–90. Such cases emphasize the magnitude of the compensatory mechanisms, and the difficulty of diagnosing the extent of haemorrhage from an examination of blood pressure and pulse rate.

FAINTING

[illegible] several times, and as this is

[illegible]

sweating.

invariably encountered, excluding injuries involving the alimentary tract and chest. These two fundamental points have been re-emphasized by the detailed examination of battle casualties and civilian casualties in air raids, by Canadian, American and British teams (Grant and Reeve, 1951; Cleghorn and Chute, 1945; Emerson and Ebert, 1945).

There was no evidence of haemoconcentration, and there was considerable evidence that blood loss alone accounted for the signs and symptoms which were observed. In the absence of strong evidence to the contrary, it can be concluded that 'shock' in cases of injury is due to blood loss.

TREATMENT

It is therefore clear that the treatment consists essentially of transfusion. Furthermore transfusion must replace the deficiency in the blood volume, i.e. restore it to the pre-injury level. This will require in many cases large volumes, i e. up to 2500 cc or six to seven bottles of citrated blood; and if transfusion is to be effective it must be given early. Animal experiments demonstrate very clearly that prolonged hypotension can lead to a condition in which complete restoration of the blood volume will not produce recovery.

It is probably worth while to re-emphasize the dangers of some methods which have been used in the past and may occasionally be employed today. Heating the patient is the worst form of treatment. Hot-water bottles and electric cradles are useless and indeed harmful. The cold, pale patient with clammy skin has either recently recovered from a vaso-vagal attack, which may have been provoked by quite a mild blood loss, or he may have had a severe haemorrhage. In the latter case his skin vessels are constricted, hence the cold skin. If he is heated then the skin vessels and the underlying muscle vessels will dilate, reducing the peripheral resistance and lowering the blood pressure. There will be an increased capacity of the cutaneous vessels and a diversion of blood from other areas. Such effects may well be disastrous. In some cases the cutaneous constriction may be so intense that vasodilatation is delayed or incomplete Local heat can then raise skin and subcutaneous

prior to examination. The striking facial pallor combined with sweating so impress that a diagnosis of severe haemorrhage may be made.

PROGNOSIS

Many of the observations mentioned in this paper have referred to moderate haemorrhage. It is frequently not realized how extensive blood loss can be in cases of injury. Battle casualties with severe wounds may lose up to 65 per cent of their total blood volume and still survive, provided that treatment is adequate.

Blood loss of over 30 per cent, or a blood volume which has been reduced to 70 per cent or less of normal, represents a critical level according to Grant and Reeve. Blood pressure is usually reduced below 100 mm Hg and heart rate increased over 100/min. when blood volume is reduced to this level. It should be realized that a blood loss of 30 per cent and a blood volume of 70 per cent are not quite the same, as haemodilution will affect total blood volume to a significant extent, at any rate after twelve hours have elapsed since time of injury. A critical level of 70 per cent blood volume, as observed by Grant and Reeve, meant that the original blood loss was usually 40 per cent or more of the initial blood volume.

Limits of the extent of blood loss compatible with survival cannot be given with precision Haemorrhages greater than 65 per cent, unless treated with immediate transfusion, probably cause death in the majority of cases. A loss of half the total blood volume, not treated with transfusion, would have a fairly high final mortality; but transfusion after twelve hours would reduce mortality and if transfusion were given within one to two hours the mortality would be quite small (Grant and Reeve, 1951).

SHOCK

It is not possible to discuss the effects of haemorrhage without some mention of the problem of 'shock' and its relationship to haemorrhage. There is a considerable degree of unanimity amongst different workers that haemorrhage, i e blood loss and hence diminution of blood volume, is the single most important factor causing eventual collapse or illness in cases of injury. It is also clear that haemodilution and not haemoconcentration is

REFERENCES

ANDERSON, D. P., ALLEN, W. J., BARCROFT, H., EDHOLM, O. G., and MANNING, G. W. (1946). *J. Physiol.* **104**, 426.

BARCROFT, J. (1934). *Features in the Architecture of Physiological Function.* Cambridge University Press.

BARCROFT, H., EDHOLM, O. G., MCMICHAEL, J., and SHARPEY-SCHAFER, E. P. (1944). *Lancet*, **1**, 489.

— and EDHOLM, O. G. (1945). *J. Physiol.* **104**, 161.

— and EDHOLM, O. G. (1946). *Lancet*, **2**, 513.

BAZETT, H. C. (1950). *Trans. 3rd Conference Factors Regulating Blood Pressure.* Josiah Macy Jr. Foundation, New York.

BEARN, A. G., BILLING, B., EDHOLM, O. G., and SHERLOCK, S. (1951). *J. Physiol.* **115**, 442.

CHUTE, A. L., CLEGHORN, R. A., and LATHE, G. A. (1945). *Proc. 8th Meeting Assoc. Committee Army Medical Research*, Vol. 2. Reports of No. 1 Research Unit, Ottawa. C.6277, II.

CLEGHORN, R. A., and CHUTE, A. L. (1945) *Proc. 8th Meeting Assoc. Committee Army Medical Research*, Vol. 2. Reports of No 1 Research Unit, Ottawa. C.6277, I.

COURNAND, A., RILEY, R. L., BRADLEY, S. E., BREED, E. S., NOBLE, R. P., LAWSON, H. D, GREGERSEN, M I., and RICHARDS, D W. (1943). *Surgery*, **13**, 964.

COURTICE, F. C., and GUNTON, R. W. (1949). *J Physiol.* **108**, 418

DE WARDENER, H. E, and MCSWINEY, R R. (1951). *Clin Sci.* **10**, 209

EDHOLM, O. G. (1949) *Fed. Proc.* **8**, 39.

EMERSON, C. P., and EBERT, R. V. (1945). *Ann. Surg* **122**, 745.

GIBSON, J. G., SELIGMAN, A. M., PEACOCK, W C., AUB, J C., FINE, J., and EVANS, R D. (1946). *J clin. Invest* **25**, 625, 848.

GRANT, R. T., and REEVE, E. B (1951). *Medical Research Council Special*

A. S

WIGGERS, H. C., and INGRAHAM, R. C. (1946). *J. clin. Invest.* **25**, 30

temperature higher than will be the case when the local blood flow is increased. So heating will either provoke vasodilatation and may precipitate vascular collapse or cause localized thermal injury.

The dangers of morphia should also be well known. In the event of low blood pressure and peripheral vasoconstriction, subcutaneous injections of morphia will be absorbed extremely slowly. Therefore there is a considerable risk that several injections may be given to control pain. If the blood pressure recovers and the cutaneous and subcutaneous blood flow increases there will be a rapid absorption of the remaining morphia. Animal work on the effects of haemorrhage shows that morphia given in relatively small doses increases the mortality rate (Gregersen, 1948).

SUMMARY

The cardiovascular effects of haemorrhage on man can be distinguished in certain respects from the results obtained in animals. There is no clear evidence of the existence of blood reservoirs of any significance in man. It is probable that the large veins contract considerably after haemorrhage, so taking up the reductions in blood volume Haemodilution is slow, on the average complete restoration of the original blood volume after 20 per cent haemorrhage takes twenty-four hours. Haemorrhages of a litre reduce cardiac output by approximately 20 per cent, but blood pressure and pulse rate remain virtually unchanged.

Peripheral vasoconstriction occurs principally in the splanchnic area, including the renal vessels, and possibly in the skin. The muscle vessels do not constrict with moderate haemorrhage.

Vaso-vagal attacks increase in frequency as haemorrhage becomes larger, and there is approximately 50 per cent incidence with blood loss of one litre. Although blood pressure recovers rapidly, facial pallor and a slow pulse rate persist for about an hour after the attack.

Haemorrhage associated with traumatic injury is always followed by haemodilution and not haemoconcentration

The treatment for haemorrhage is early and adequate blood transfusion.

systolic and diastolic pressure; 'reversibility' implies the possibility of producing a more lasting lowering of raised blood pressure or even of restoring it to normal.

that, as long ago as 1900, Gumprecht reported that the level of blood pressure varied with rest and with emotion. The succeeding half-century has seen persistent efforts to explain not only the occurrence of hypertension in man, but also its surprising lability.

The realization that certain forms of human hypertension, even in the clinically malignant phase, are reversible marks an important advance in our knowledge of hypertension. The removal of a diseased kidney may, within a few minutes, restore to normal levels a blood pressure that has been raised for years, while the successful removal of a phaeochromocytoma is accompanied by a similar dramatic reversal. The newer methods of treatment of hypertension, by surgical or medical sympathectomy, may also dramatically reduce the peripheral resistance and blood pressure and, in so doing, may relieve the manifestations of malignant hypertension. Cournand (1949) has studied five cases of malignant hypertension before and after sympathectomy. The cardiac output was originally within normal limits and remained unchanged post-operatively; the mean arterial pressure and hence the total peripheral resistance was significantly decreased.

When we consider the causes for this reversible increase in arteriolar tone, it is clear that at least two possibilities exist. There may be an increased neurogenic vasoconstriction or the intrinsic tone of the vessels may be increased, possibly as a result of circulating chemical stimulants. Evidence suggests that both neurogenic and humoral factors are of importance.

Investigation of the control of blood pressure in animals has

from the aortic arch and carotid sinus, the chemically stimulated

VI

Lability of Blood Pressure

M. L. ROSENHEIM

In a series of lectures devoted to the scientific basis of medicine, it is clearly presumptuous to discuss such a subject as the lability of blood pressure, for, while there has been an enormous volume of experimental work on animals and clinical observation and experiment on man, we have scarcely got be- [illegible] sure. In most hypertensive patients, at least in the early stages, the blood volume, the viscosity of the blood and the elasticity of the walls of the great vessels are normal, and the actual level of the blood pressure and changes in that level can be roughly regarded as the resultant of two opposing forces; cardiac output and the peripheral resistance.

The cardiac output has been repeatedly measured in patients with both essential and chronic renal hypertension and found to be normal, and it is now generally accepted that the underlying abnormality is a generalized increase in peripheral vascular resistance. This increased resistance might be due either to an increase in arteriolar tonus or to organic changes in the arteriolar walls. Histological studies demonstrate that such anatomical changes as do occur in the vessels do not precede the occurrence of the hypertension. This view is supported by experimental work on animals and by observation of the lability and reversibility of raised blood pressure in man.

The term 'lability' is applied to those transient fluctuations, both upwards and downwards, that can be observed in both the

THE CEREBRAL CIRCULATION AND HYPERTENSION

A further important advance in our knowledge of the circulation in hypertension was made when Kety and his colleagues (Kety, Hafkenschiel, Jeffers, Leopold and Shenkin, 1948) showed that the cerebral blood flow in the uncomplicated hypertensive patient was normal. This had been roughly demonstrated in the past (cf. Williams and Lennox, 1939), but the new nitrous oxide method allowed more precise measurement and the calculation of the cerebral vascular resistance.

It is easy to state, as has so often been done, that the raised arterial blood pressure is the result of increased arteriolar tonus, but this simple statement leaves two important facts unexplained. First, why does the rise of blood pressure not call into play the various compensatory reflexes, especially since such reflexes have been shown to be active in the hypertensive? Secondly, why does the blood pressure rise to a level exactly adequate to maintain normal cerebral blood flow?

The control of blood flow through the tissues is highly complex and varies from organ to organ. Local regulatory mechanisms for maintaining blood flow have been demonstrated in muscle, where ischaemia produces local vasodilatation, either by axon reflex or direct action of metabolites on the vessels. The renal blood flow, as measured by clearance techniques, is diminished in many patients with essential hypertension and it has been argued that this implicates a renal mechanism in human hypertension. The renal vasoconstriction is, however, also reversible and is probably part of the generalized increase in arteriolar tone. It would appear that the kidney possesses no local regulatory mechanism to compensate for increased vascular resistance or ischaemia.

At least the [illegible]

stimulation of the vasomotor centre with resultant neurogenic vasomotor reflexes affecting the arterial pressure. The tone of the cerebral vessels must be increased in hypertension. If it were not, the rise in blood pressure resulting from the increased peri-

reflexes from the carotid body, postural vasoconstrictor reflexes and those associated with change in cardiac output or venous return. Many of these reflexes can be demonstrated to follow acute change in man. Thus the sudden rise of blood pressure due to increased peripheral resistance following the intravenous infusion of noradrenaline is accompanied by bradycardia, which is abolished by atropine (Swan, 1949).

The problem of chronic hypertension is, of course, different, for in both human hypertension and experimental renal hypertension in animals, the rise of blood pressure follows a general increase in peripheral vascular tone and the rise in pressure does not apparently stimulate the depressor reflexes which respond to acute changes. This failure might suggest a breakdown of the normal reflex mechanisms, but Pickering, Kissin and Rothschild (1936) have investigated the carotid sinus reflex in hypertensive patients and found it active. Human hypertension does not resemble that produced in animals by denervation of both carotid sinuses. Incidentally, treatment of the carotid sinus syndrome in man, by bilateral denervation, does not lead to hypertension (Turner and Learmonth, 1948). It is possible in fact to show that, at higher levels of blood pressure, reflex vasomotor control is still active and persistent arterial hypertension in man cannot be explained in terms of failure of neurogenic reflexes.

I must digress for a moment at this point to mention one further reflex arc, brought into prominence by the recent renewal of interest in the veratrum alkaloids As long ago as 1867, Bezold and Hirt described the hypotensive effect of veratrum and showed that it evoked a vagal reflex resulting in bradycardia. This effect has been intensively studied by Dawes, Mott and Widdicombe (1951a, b). They have demonstrated that the reflex arises from receptors in the coronary circulation and that it is abolished by cooling the vagi. The first effect is a reflex bradycardia followed, with larger doses, by peripheral vasodilatation. There is also evidence that some of the alkaloids have a central action and this may be the main site of action in man. Neither normal function nor the stimulus of this reflex is clear.

pheral vasoconstriction, probably as a result of medullary ischaemia. I would draw attention to the importance of normal cerebral blood flow and to the fact that here again, while compensation occurs, the rise in blood pressure maintains a normal cerebral flow.

The cerebral flow can, however, undergo adjustment without evoking reflex vascular change. Kety and his colleagues (Hafkenschiel, Shenkin, Kety and Jeffers, 1949) found the cerebral blood flow and cerebral oxygen metabolism unchanged after bilateral thoraco-lumbar splanchnicectomy for hypertension. Cerebral vascular resistance was lowered, the cerebral vessels apparently having undergone compensatory dilatation following the fall in arterial blood pressure. Kety concluded that a prolonged or moderate fall in the blood pressure of hypertensives was not associated with the expected reduction in cerebral blood flow, when measured in the supine position. The vessels of the brain appeared capable of compensatory dilatation. It is clear that the reciprocal balance between cerebral vascular and peripheral resistance is complex.

Striking changes in peripheral vascular tone accompany a change from supine to erect posture and, in both normal and hypertensive subjects, little change in blood pressure occurs. The cardiac output falls, probably as a result of diminished venous return, and tachycardia and reflex vasoconstriction occur (McMichael and Sharpey-Schafer, 1944). These reflexes function in the hypertensive. Anyone who has seen the dramatic cerebral symptoms that result when, following ganglionic block-[illegible]

the autonomic nervous system.

When we look at the various mechanisms subserving the maintenance of blood pressure in both animals and man, it is clear that these can be regarded as an intricate system for maintaining adequate cerebral circulation and oxygenation. The carotid sinus, initiating compensatory vascular and cardiac changes in response to sudden change in blood pressure, and the carotid body, sensitive to chemical change, both lie on the main

pheral resistance elsewhere would lead to an increased cerebral blood flow at the expense of other tissues. The maintenance of a normal cerebral blood flow in the face of increased arteriolar resistance appears to me to be one of the major homeostatic triumphs of the body and to merit further discussion.

While sympathetic fibres to the cerebral vessels have been demonstrated and while stimulation of the vagus and carotid sinus nerves has been shown in animals to produce dilatation of the pial vessels (Forbes, Mason and Wortman, 1937), neurogenic control is relatively unimportant compared to chemical control and is far less active than the neurogenic control of the peripheral circulation (Forbes and Cobb, 1938).

Kety and his colleagues have studied the cerebral circulation under various circumstances and have demonstrated a variety of controlling mechanisms. They have confirmed and extended the findings of earlier workers, that the regulation of the cerebral blood flow appears to depend largely upon changes in the carbon dioxide and oxygen tension in the arterial blood.

Kety and Schmidt (1948) have shown that, in normal man, an increased arterial tension of carbon dioxide causes a marked increase in cerebral blood flow. This results from peripheral vasoconstriction with a rise in arterial pressure, the cardiac output remaining unaltered. Anoxia also leads to an increased cerebral blood flow, though here an increased cardiac output played a large part. In the brain chemical changes can thus apparently initiate widespread reflex vascular changes through the vasomotor centre.

In a further paper Kety and his colleagues (Kety, Shenkin and Schmidt, 1948) showed, from a study of patients with cerebral tumours, that as the intracranial pressure rose, cerebral vascular resistance passively increased and that up to a cerebrospinal pressure of some 450 mm of water, there was a compensatory rise in arterial pressure, and, up to this pressure, cerebral blood flow was fairly well maintained The patients who failed to compensate and who exhibited a marked reduction in cerebral blood flow were all comatose. These findings confirmed Cushing's original observation (1901) in animals that an acute rise in c.s.f. pressure led to a rise in arterial pressure by peri-

hypertensive patient, many of them aiming at the measurement of the basal pressure, others at producing a rise from the base line.

In the early stages of hypertension, whether due to chronic renal disease or essential hypertension, the blood pressure falls to normal with rest. Casual readings are raised, basal are normal. This is the earliest clinical stage of hypertension. The large-scale investigation of blood pressure in United States Army officers (Levy, Hill, Stroud and White, 1944; Levy, White, Stroud and Hillman, 1945) suggested that a single raised reading early in life is associated with a significantly greater invalidism from hypertension in later years, and several writers have shown that transient hypertensive casual readings may precede the development of true hypertension (Diehl and Hesdorfer, 1933; Hines, 1940).

As the condition progresses, the blood pressure fails to return to normal with rest; the basal pressure is now raised. Campbell and Blankenhorn (1925) showed that patients whose blood pressure fell to normal with sleep had less evidence of disease and less disability than those whose blood pressure remained raised in sleep. It is probable, but difficult to prove, that it is the height of the basal pressure rather than the casual, the persistent stress on the arterial and arteriolar walls, that is responsible for the secondary anatomical changes that we recognize in the later stages of the benign and in the malignant phase of hypertension.

Can a prehypertensive phase be detected? Is there a stage in hypertensive vascular disease when even casual readings of the blood pressure are normal, but in which pathological over-reaction to blood pressure raising tests occurs?

Various authors have claimed that such a stage exists and, among others, the cold pressor test and breath-holding test

s means

reactors

that, in

their family histories, an abnormally high percentage of hypertensive parents will be found, have not been fully confirmed or generally accepted, nor is every hypertensive patient a hyper-reactor (see Wolff, 1951). The cold pressor test measures a

pathway of the cerebral circulation. The cerebral vessels do not participate in the general reflex vasoconstriction to the same extent that do the peripheral vessels and the gaseous tension of the blood flowing through the brain can initiate changes in peripheral vascular tone and cardiac output.

It appears possible that, in chronic hypertension in man, the maintenance of a normal cerebral blood flow takes precedence over the expected depressor vagal reflexes and that, as far as the cerebral circulation is concerned, the rise of blood pressure may be beneficial. Whether the initial rise in general arteriolar tone is neurogenic or humoral, the cerebral homeostatic mechanism may permit the rise of blood pressure to a new level at which cerebral blood flow is normal and, at this level, the usual vasomotor reflexes are active. This is pure hypothesis.

THE LABILITY OF BLOOD PRESSURE

Even in a normotensive subject the arterial blood pressure can readily be made to vary; it will rise with certain visceral stimuli, with exercise, pain and emotion, and will settle with rest or sleep. *One of the striking features of the blood pressure in hypertensive subjects is the greater tendency for such swings to occur.* It has long been known that raised blood pressure often falls during sleep or complete mental and physical rest, and the tendency for the pressure to fall during hospitalization has proved a major difficulty in assessing the efficacy of new lines of treatment.

In 1919 Tixier introduced the term 'tension residuelle', realizing that under basal conditions the blood pressure of hypertensives reached a lower, though not necessarily normal, level, and that various activities and stimuli usually kept it above this basal level. In 1943, Smirk and his colleagues (Alam and Smirk, 1943; Gatman, Amin and Smirk, 1943, Smirk, 1944) stressed the importance of measuring this 'basal pressure'. They regarded the casual pressure, the chance reading as composed of this basal pressure with a superadded or 'supplemental' element. The basal pressure was shown by Kilpatrick (1948) *to be more consistent than casual readings.*

Various tests have been devised for the investigation of the

Ackerman, Cohn, Schroeder and Steel, 1945). Given an increased reactivity, an increased arteriolar tone, it is conceivable that recurrent emotional stress, whether conscious or subconscious, could lead to progressive deterioration. Does the subconscious hostility or the increased arteriolar tone come first?

BASAL BLOOD PRESSURE

Various methods have been adopted for the measurement of the basal blood pressure. Complete physical and mental rest, though difficult to achieve, is as reliable as any. Ayman and Goldshine (1940) have stressed the difference between blood pressure readings taken at hospital by the doctor and those taken by a relative in the patient's own home.

Cady, Horton and Adson (1936) introduced the use of barbiturate sedation into the study of hypertension, with the amytal test. Hammarström (1947) showed that the lowest systolic and diastolic pressures recorded under the influence of amytal were practically identical with the lowest readings obtained during twenty-four hours' quiet existence.

We have used a rapidly acting barbiturate 'Seconal' for the sedation test and found that from the casual diastolic reading it was impossible to predict the basal level that would be reached with sedation (Frew and Rosenheim, 1949).

With the introduction of the ganglion-blocking drugs, it was hoped that it might be possible to remove the supplemental neurogenic tone in hypertensive patients, leaving undisturbed the basal pressure.

Acheson and Moe (1946) showed that the action of the tetra-ethyl-ammonium ion was to block the transmission of autonomic impulses at both sympathetic and parasympathetic ganglia. It was soon found that tetra-ethyl-ammonium salts produced a marked fall in the blood pressure of hypertensive patients In normals little or no fall occurred In hypertensives the amount of fall varies and it is impossible to predict, from the casual level, what fall will occur. The lowering of blood pressure is due to the interruption of vasoconstrictor impulses. The cardiac output is unaltered or slightly increased, the peripheral resistance markedly lowered (Lyons, Hoobler, Neligh, Moe and

response to pain and is of interest in that it demonstrates the neurogenic reflex rise in blood pressure—the effect being reduced or abolished by ganglion-blocking drugs. It is of course to be expected that any neurogenic reflex would have a greater blood pressure raising effect if the arteriolar tone were already increased. The prehypertensive state must exist but has, so far, not been indisputably demonstrated.

Much has been written about the effect of emotion on blood pressure (Alexander, 1939). This can be considered from two different aspects. In many patients with essential hypertension there is an abnormal hypertensive response to emotional stress. The blood pressure of a normal person readily jumps above the basal level if he is disturbed or irritated. In the hypertensive, discussion of certain subjects produces a sharp rise in pressure and, in the study of the psychiatric make-up of patients, these emotionally charged topics appear of aetiological significance. It has been shown that in some cases the rise of pressure is due to sudden increase in cardiac output with insufficient compensatory vasodilatation, in others neurogenic vasoconstriction occurs (Wolf, Pfeiffer, Ripley, Winter and Wolff, 1948; Wolf and Wolff, 1951). It is, once more, not difficult to appreciate how such sudden changes, whether in cardiac output or vasoconstriction, might produce a greater response in patients with raised arteriolar tone.

Psychiatrists, however, go further and claim to recognize in hypertensive patients a somewhat uniform underlying reaction or personality. The typical hypertensive has, they claim, had a difficult childhood, has made an incomplete adjustment and has been forced to stifle his aggressive tendencies. He reaches adult life filled with aggressive drive that is inhibited and not overt and with subconscious feelings of hostility and anxiety. The evidence for a psychosomatic factor in hypertension is convincing, but it is difficult to reconcile the claim that essential hypertension is entirely psychogenic with the established fact that it is hereditary. Further it must be noted that this typical picture of buried aggression has been found in patients with hypertension associated with hydronephrosis, pyelonephritis and congenital renal aplasia as well as with essential hypertension (Binger,

T.E.A. is probably purely ganglionic, a maximal result is not always obtained. The exact site of the hypotensive action of barbiturates is not certain and a lowered cardiac output has been reported following amytal (Winchell, Taylor and Chapman, 1951).

Hexamethonium has a much more powerful and more prolonged action than T.E.A. It has a similar ganglion-blocking action, but is not believed to stimulate the secretion of adrenaline or noradrenaline. Hexamethonium will reduce the blood pressure substantially below the basal level as measured by T.E.A. or barbiturates. There is some evidence that the cardiac output may be lowered as a result of venous pooling (Werko, Frisk, Wade and Bliasch, 1951), the action being more prolonged, and, unless the patient lies quite flat part of the lowering is certainly postural in origin The beneficial results obtained in the treatment of hypertension by both thoraco-lumbar splanchnicectomy and the methonium drugs while partly due to the direct decrease in peripheral resistance must owe much to the postural element which is often prominent in successful cases

Studies with T.E.A. and hexamethonium leave little doubt that neurogenic impulses play a large part in the production of [illegible]

those with essential hypertension. This may be true in the severer cases, but in my experience lability is marked in the early stages of renal hypotension, and, even in the malignant phase, chronic renal hypertension can be dramatically improved by medical or surgical sympathectomy.

The hypertension of acute nephritis is very different to that of chronic renal disease, for it has been shown that during the early stage of this disease the raised blood pressure is fixed and does not fall either with sedation or T E.A., though it may drop to normal and be normally labile a few days later (Ferris, Reiser, Stead and Brust, 1948; Frew and Rosenheim, 1949). The hypertension of acute nephritis is apparently not neurogenic and it has been suggested that it may be due to a renal humoral factor or be related to the marked retention of sodium (Rosenheim, 1951).

Peet, 1948). The effect is counteracted by adrenaline, which acts distally to the block and, in some patients, maximal falls may not be obtained, the drug apparently stimulating the adrenal medulla.

In 1949 we were able to show that there was a significant correlation between the basal diastolic pressure obtained with T.E A. and that obtained with a rapidly acting barbiturate (Seconal) (Frew and Rosenheim, 1949). A similar correlation was reported between T.E.A. and amytal (Tamagna and Poindexter, 1948) and, for young adults with hypertension between T.E.A. and the spontaneous minimal blood pressure (Frisk, Hammarstrom, Lagerlof, Werko, Bjorkenheim, Holmgren and Larsson, 1948). At that time, it appeared that complete rest, barbiturate sedation and T.E.A. would all produce the same basal pressure and that the supplemental or neurogenic element was temporarily removed.

It is of interest to note, in passing, that in the series of hypertensive patients whose basal pressure we investigated the five patients with papilloedema, i.e. in the malignant phase of hypertension, had a high basal level with both T.E.A. and barbiturate, and Lyons and his colleagues (Lyons, Hoobler, Neligh, Moe and Peet, 1948) found that patients who failed to have a drop of 10 per cent or more from their initial diastolic pressure tended to have greater retinal changes, a larger heart and more impaired renal function. This suggests again that secondary anatomical changes are related to the height of the basal pressure. It is impossible to predict from the results obtained either with T.E.A. or with amytal which patients will respond well to thoraco-lumbar splanchnicectomy. The basal blood pressure is not a sound guide to surgical prognosis.

If it were possible clearly to split raised blood pressure into a basal or humoral and a supplemental or neurogenic element, further study would be greatly simplified. Unfortunately it is probably not. Firstly the basal pressure, though more constant than the casual, is itself somewhat variable from day to day and Ferris and his colleagues (Ferris, Reiser, Stead and Brust, 1948) considered that there was a reciprocal interplay between the humoral and neurogenic elements. Secondly while the effect of

thoraco-lumbar splanchnicectomy. I have no personal experience of these tumours and little has been recorded about the lability of the blood pressure. Smithwick (Smithwick, Greer, Robertson and Williams, 1950) states that the cold pressor test is normal, others (Evans, Rubitsky, Bartels and Bartels, 1951) have found it increased. Swan (1951) has reported one in which amytal produced little fall of blood pressure and T.E.A., so far from producing a fall, may precipitate a hypertensive crisis and has been advocated as a test for the presence of a tumour (La Due, Murison and Pack, 1947). It is interesting to note that one case has been reported in which a low sodium diet produced a fall in pressure and relief of papilloedema (Kositchek and Rabwin, 1950).

The symptoms of a phaeochromocytoma are almost certainly due to the liberation of adrenaline and noradrenaline into the blood stream. These substances have been isolated from the urine of patients and detected in large amounts in the tumours. Circulating adrenaline cannot, of course, mimic essential hypertension, for it constricts the skin vessels, causes muscular vasodilatation and increased cardiac output. Noradrenaline, on the other hand, on slow intravenous infusion, produces a state closely resembling human hypertension (Goldenberg, Pines, Baldwin, Greene and Roh, 1948). There is an increase in peripheral resistance with no alteration in cardiac output.

Can noradrenaline and adrenaline play any part in essential hypertension? Studies with the newer adrenergic blocking agents suggest that this is unlikely. A variety of drugs have recently been introduced which antagonize the vasoconstrictor action of circulating adrenaline and noradrenaline; apparently blocking the response of the effector cells to the action of these compounds. Two have been widely studied in hypertension, dibenamine and benzodioxane, others, Priscol, Regitin (C.7337) and dihydroergocornine, have been chiefly investigated in peripheral vascular disease.

Dibenamine, dibenzyl beta-chloroethylamine, is a most potent

The present view, based on the action of the ganglion-blocking drugs, that there is a large neurogenic element in essential hypertension and none in the hypertension of acute glomerular nephritis, is directly opposed to the findings of Pickering (1936) who, as a result of measuring the blood flow through the hand before and during full vasodilatation produced by warmth, concluded that the increased peripheral resistance in essential [illegible]

bably humoral, his experiments, especially as regards acute nephritis, require repetition using other means of producing vasodilatation.

How far is neurogenic control a factor in maintaining normal blood pressure? Heymans (1950), from his animal experiments, considered that 'the vasomotor centre was continuously active, *sending out a stream of constrictor impulses . . . to the peripheral blood vessels*'. If a large dose, 100 mgm, of hexamethonium bromide is injected intravenously in a supine normotensive man, a small fall of blood pressure may occur, but in some the blood pressure remains practically unaltered. That neurogenic control has been removed can readily be demonstrated by standing the subject up, and full vasodilatation can also be shown. The blood pressure in such persons, while lying flat, must be maintained by the intrinsic tone of the arterioles or by their humoral stimulation.

HUMORAL MECHANISMS IN HYPERTENSION

Many possible humoral mechanisms for the maintenance of normal or increased peripheral resistance have been investigated; four require special consideration.

One well-established form of humoral hypertension is that associated with a suprarenal medullary tumour, the phaeochromocytoma. It was originally believed that these tumours produced only intermittent crises of hypertension, but it is now clear that persistent hypertension may occur and that this can closely mimic essential or chronic renal hypertension, some cases presenting in the malignant phase of hypertension. Some unsuspected tumours have been found in patients selected for

(Raab, Humphreys and Lepeschkin, 1950) found that deoxycortone increased the pressor effect of adrenaline and noradrenaline in man. Hypertension occurs in Cushing's syndrome and has been observed in many patients under treatment with cortisone or ACTH. The importance of the adrenals in hypertension has recently been emphasized by the important work being done at the London Hospital (Floyer, 1951; Ledingham, 1951).

There appears to be synergistic action between the adrenal cortical hormones and sodium. The effect of low sodium diets on hypertension, first suggested many years ago, has recently roused great interest again. Therapeutic sodium depletion produces a fall of blood pressure in the great majority of cases of hypertension. Sodium depletion appears to affect the basal blood pressure, for the blood pressure in patients on low sodium diets may fall well below the basal level achieved with T.E.A. (Stead, Reiser, Rapaport and Ferris, 1948) or sedation prior to treatment, and there is some evidence that the addition of sodium to depleted patients or those with Addison's disease raises the basal and not the supplemental or neurogenic pressure (Rosenheim, 1951). Sodium and the adrenal cortical hormones are humoral agents that may play an important role in certain phases of hypertension in man and Perera and Blood (1947) have shown that patients with uncomplicated essential hypertension react differently to normal man on sudden sodium withdrawal. They do not lose so much weight, apparently holding more tightly to their sodium. Dr. A. G. Spencer working in the Medical Unit at University College Hospital has shown that the hypertensive response to sodium depletion can be mimicked in the normal by the administration of deoxycortone Thomas (1951) has claimed to have found a similar reaction to sodium deprivation in the prehypertensive phase, but this awaits confirmation

I have so far avoided reference to the possible renal origin of hypertension and the vast amount of work that has been done on experimental hypertension since Goldblatt's original papers. The renin/angiotonin story is fascinating, and renal ischaemia in animals produces a raised blood pressure that in many ways

the hypertension due to released adrenaline or noradrenaline. Given by slow intravenous drip, it produces minimal changes in normal blood pressure, though orthostatic hypertension occurs (Hecht and Anderson, 1947). In patients with severe hypertension, a marked fall in blood pressure results (Wunsch, Warnke and Myers, 1950). The toxic effects of this drug prevent its more widespread use.

Benzodioxane has also been widely used as a test for the presence of a phaeochromocytoma, though its value for this purpose has recently been doubted. In the typical case, the intravenous infusion of benzodioxane leads to a fall of blood pressure, the drug counteracting the direct vasoconstrictor action of circulating adrenaline or noradrenaline. In essential hypertension, however, no such fall is produced and, in many cases, a somewhat alarming sharp rise of blood pressure occurs. Nickerson (1951) concludes that benzodioxane blocks responses to circulating adrenaline much more readily than those due to sympathetic nerve activity and it has been suggested that the pressor action of the drug is due to central stimulation or to an increase in cardiac output. The fact that these two drugs are used to distinguish the hypertension of a phaeochromocytoma from that of essential hypertension does suggest that circulating adrenal medullary hormones play little or no part in essential hypertension. It may be noted that none of the specific adrenergic blocking agents inhibits the vascular response to angiotonin (Nickerson, 1951).

The adrenal cortex is also closely associated with hypertension, and the relation between the various adrenal cortical hormones, sodium and hypertension is gradually being unravelled. It has long been known that experimental hypertension cannot exist in the absence of adrenals, and that deoxycortone will raise the lowered blood pressure of a patient with Addison's disease. The hypertensive patient is especially sensitive to deoxycortone and has a larger hypertensive response than a normal person. It has been shown that this response is not due to an increase in blood volume (Perera and Blood, 1947), nor to an increased cardiac output (Goldman and Schroeder, 1948) and is, presumably, the result of direct action on the arterioles. Raab and his colleagues

depend on increased arteriolar tone, it is not surprising to find many similarities between them, but these similarities do not necessarily imply similar underlying mechanisms.

I have here tried to marshal some of the better established facts concerning hypertension in man and its lability. It is clear that the problem is being approached along many and divers lines and that some of these lines are converging. There are many contributory factors in any single case of hypertension, and while each single factor can perhaps be best studied by animal experiment, the problem of hypertension as a whole must be tackled, and ultimately solved, by study in man.

REFERENCES

ACHESON, G. H, and MOE, G. K. (1946) *J. Pharmacol* **87**, 220.
ALAM, G. M., and SMIRK, F. H. (1938) *Clin Sci.* **3**, 259.
— (1943) *Brit. Heart J* **5**, 152, 156.
ALEXANDER, F. (1939) *Psychosom. Med* **1**, 173
AYMAN, D, and GOLDSHINE, A. D. (1939) *Arch. intern. Med* **63**, 899.
— and GOLDSHINE, A. D. (1940) *Am J med. Sci* **200**, 465.
BEZOLD, A VON, and HIRT, L (1867). *Untersuch physiol Lab Würzburg*, **1**, 73
BINGER, C A. L, ACKERMAN, N W, COHN, A E., SCHROEDER, H. A, and STEELE, J. M. (1945) *Psychosomatic Monographs*. New York.
CADY, J. B., HORTON, B T, and ADSON A. W. (1936). *Proc. Mayo Clin.* **11**, 825
CAMPBELL, H. E, and BLANKENHORN, M A (1925). *Am. Heart J.* **1**, 151.
COURNAND, A. (1949) *Factors Regulating Blood Pressure* Josiah Macy Foundation, **3**, 203
CUSHING, H. (1901). *Bull J. Hopk Hosp.* **12**, 290.
DAWES, G S, MOTT, J C, and WIDDICOMBE, J G (1951a) *J. Physiol* **115**, 258
— (1951b). *Brit. J Pharmacol* **6**, 675
DIEHL, H S, and HESDORFFER, M B. (1933) *Arch intern Med* **52**, 948.
EVANS, J A, RUBITSKY, H J, BARTELS, C C, and BARTELS, E. C (1951). *Amer J Med* **11**, 448
FERRIS, E B, REISER, M F, STEAD, W W., and BRUST, A. A. (1948). *Trans. Ass. Amer. Physicians*, **61**, 95.
FLOYER, M A. (1951). *Clin Sci.* **10**, 405
FORBES, H S, NASON, G I, and WORTMAN, R. C. (1937). *Arch. Neurol.* **37**, 334

mimics the picture of essential and chronic renal hypertension in man. Yet its relevance to human essential hypertension is still non-proven. Angiotonin, injected intravenously, produces a marked increase in peripheral resistance due to increased arteriolar tone. The resultant hypertension is unaffected by T.E.A. or adrenergic blocking agents—a clear example of humoral hypertension. After a period of great enthusiasm, the pendulum has swung and the renin mechanism is no longer regarded as playing a vital part in man except in certain rare cases.

The other renal humoral mechanism that has excited interest recently is the vasomotor exciting mechanism (V.E.M.) described by Shorr and Zweifach (Shorr, Zweifach, Furchgott and Baez, 1947). They reported that an extract of anoxic kidney would sensitize the smallest arterioles of the mesentery to the direct topical action of adrenaline. The anoxic liver is said to produce another substance V.D.M. (vasodepressor mechanism), which has been identified as ferritin and which inhibits the effect of V.E.M. The interplay of these two substances has been [illegible]

The concept, however, is of great interest—that a humoral mechanism might render the arterioles more sensitive to adrenaline or, presumably, to sympathetic neurogenic impulses. We have already seen that deoxycortone has been shown to have a similar action, and Lee and Holze (1951) have recently demonstrated heightened sensitivity of the arterioles of the conjunctiva in hypertensives to topically applied adrenaline.

One undoubted fact about essential hypertension is that it is a hereditary disease—but just what is inherited? On the one hand it might be a hypersensitive vasomotor centre, a centre that responded more vigorously than normal to visceral, psychic and external stimuli; on the other hand, the primary fault might lie in the arteriolar bed which might respond more readily or vigorously to normal vasomotor impulses. Smirk (1949) has emphasized the many factors that appear to play a part in determining the incidence of essential hypertension. Since both essential and chronic renal hypertension, in the final issue,

SMITHWICK, R. H., GREER, W. E. R., ROBERTSON, C. W., and WILKINS, R. W. (1950) *New Engl. J. Med.* **242**, 252.
STEAD, W. W., REISER, M. F., RAPAPORT, S, and FERRIS, E. B. (1948). *J Clin. Invest* **27**, 766.
SWAN, H. J. C (1949). *Lancet*, **2**, 508.
— (1951) *Brit. med. J.* **1**, 440.
TAMAGNA, I. G., and POINDEXTER, C. A (1948) *Amer. J. med. Sci.* **215**, 651.
THOMAS, C. B (1951). *Bull. J. Hopk. Hosp.* 89, 419.
[illegible]
[illegible]
[illegible]
[illegible]
WINCHELL, P., TAYLOR, H L, and CHAPMAN, C. B. (1951). *Circulation*, **4**, 228.
WOLF, S., PFEIFFER, J. B, RIPLEY, H S, WINTER, O. S, and WOLFF, H. G. (1948). *Ann. intern. Med.* **29**, 1056.
— and WOLFF, H. G. (1951). *Hypertension A Symposium.* Univ of Minnesota Press, Minneapolis, 288
WOLFF, H. H. (1951). *Quart. J. Med.* **20**, 261.
WUNSCH, R. E, WARNKE, R. D, and MYERS, G. B (1950) *Ann. intern. Med.* **33**, 613.

[illegible] 18
[illegible] ENHEIM,
[illegible]
G [illegible]
G [illegible]
(1948). *Amer. J. Med.* 5, 792.
GOLDMAN, M. L., and SCHROEDER, H. A. (1948). *Amer. J. Med.* 5, 33
GUMPRECHT (1900). *Ztschr. klin. Med.* 39, 377.
HAFKENSCHIEL, J. H., SHENKIN, H. A., KETY, S. S., and JEFFERS, W. A. (1949). *J. clin. Invest.* 28, 786.
[illegible]
HINES, E. A. (1940). *J. Amer. med. Ass.* 115, 271.
KETY, S. S., HAFKENSCHIEL, J. H., JEFFERS, W. A., LEOPOLD, I. H., and SHENKIN, H. A. (1948). *J. clin. Invest.* 27, 511.
— and SCHMIDT, C. F. (1948). *J. clin. Invest.* 27, 484.
— SHENKIN, H. A., and SCHMIDT, C. F. (1948). *J. clin. Invest.* 27, 493.
KILPATRICK, J. A. (1948). *Brit. Heart J.* 10, 48
[illegible] 806
[illegible] 118
[illegible] 944).
J. Amer. med. Ass. 126, 829.
— WHITE, P. D., STROUD, W. D., and HILLMAN, C. C. (1945). *J. Amer. med. Ass.* 128, 1059
LYONS, R. H., HOOBLER, S. W., NELIGH, R. B., MOE, G. K., and PEET, M. M. (1948). *J. Amer. med. Ass.* 136, 608.
McMICHAEL, J., and SHARPEY-SCHAFER, E. P. (1944). *Brit. Heart J.* 6, 33.
NICKERSON, M. (1951). *Hypertension. A Symposium.* Univ. of Minnesota Press, Minneapolis, 410.
PERERA, G. A., and BLOOD, D. W. (1946). *Amer. J. Med.* 1, 602.
— and BLOOD, D. W. (1947). *Ann. Int. Med.* 27, 401.
PICKERING, G. W. (1936). *Clin. Sci.* 2, 209 and 363
— KISSIN, M., and ROTHSCHILD, P. (1936). *Clin. Sci.* 2, 193
RAAB, W., HUMPHREYS, R. J., and LEPESCHKIN, E. (1950). *J. clin. Invest.* 29, 1397
ROSENHEIM, M. L. (1951). *Lancet*, 2, 505.
SHORR, E., ZWEIFACH, B. W., FURCHGOTT, R. F., and BAEZ, S. (1947). *Factors Regulating Blood Pressure.* Josiah Macy Found. 1, 32
SMIRK, F. H. (1944). *Brit. Heart J.* 6, 176.
— (1949). *Brit. med. J.* 1, 791

illumination; the abscissae are the wavelengths of light with blue on the left and red on the right and the ordinates are the relative sensitivity of the eye. Curve (a) shows the sensitivity of

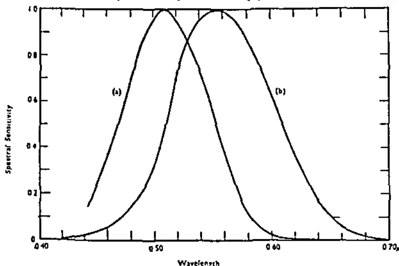


FIG 1. (After Wright, 1946.) The spectral sensitivity of the dark-adapted (a) and light-adapted (b) eye. Each curve is plotted as a percentage of its own maximum.

the dark-adapted or scotopic eye and curve (b) that of the light-adapted or photopic eye. The point of maximum sensitivity changes on light adaptation from wavelength 500 mμ to wavelength 555 mμ and the presence of two separate sensitivity curves is the best evidence we have for the dual nature of the visual system.

THE DARK-ADAPTED EYE

Let us consider the dark-adapted or scotopic eye first. The eye in this condition is being used as a simple detector of light; form vision is poor; the appreciation of colour is not possible so that the sensations experienced are merely those of black and white, and the system appears to be functioning as though it had but a single mechanism. Evidence in support of this is derived from the measurement of the absorption curve of the photosensitive pigment, visual purple, and the comparison of this curve with the sensitivity curve in Fig. 1.

VII

The Physiological Basis of Visual Sensation

L. C. THOMSON

In choosing the title for this lecture I thought it would give me an opportunity to discuss the mechanisms in the human eye. Recently there has been considerable discussion amongst physiologists as to the number of systems present in the visual pathway, and my purpose here is to examine the evidence to see whether a decision on this question is possible.

Anatomically the visual system is divided into a number of parts. The light at first passes through the dioptric mechanism of the eye to form images of external objects on the retina. These images are then recorded by a process which takes place in cells within the retina called photoreceptors and consists of the absorption of light by a photosensitive pigment which changes its nature and initiates a nervous discharge in the subsequent neurones of the retina. This nervous discharge forms the message which is sent over the transmission system of nerves to the brain. Here in the brain the message gives rise to a sensation of light or colour and can also be used to initiate motor reactions elsewhere in the body.

The dual nature of the visual system has been known since the end of the last century. We possess two separate receiving mechanisms intermingled in the eye and optic nerve; one, the scotopic mechanism used at low luminance at night and the other, the photopic system for daylight vision and the appreciation of colour.

Fig. 1 shows the relative sensitivity of the eye at two levels of

that the spectral sensitivity curve of the dark-adapted eye would be similar in shape to the absorption spectrum of visual purple over the visible region of the spectrum, because at those wavelengths at which absorption is greatest there should clearly be the greatest sensitivity. Comparisons of this kind have been made a number of times during the last 100 years and one such is shown in Fig. 3.

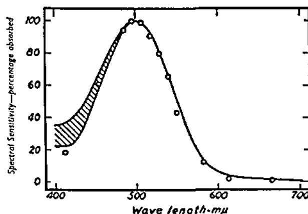


FIG 3 (After Wald, 1938.) A comparison between the spectral sensitivity of the dark-adapted eye (points) and the absorption spectrum of visual purple (full line and hatched area) The sets of data have been equated at the maximum

In this diagram the points are the spectral sensitivity curve of the dark-adapted eye and the hatched area the limits within which the absorption spectrum of visual purple lies The agreement between these two sets of data is so satisfactory that one can say that the dark-adapted eye uses a single mechanism and that that mechanism relies on the single pigment visual purple. The agreement between these two sets of data can be made even more perfect by making several minor corrections to the data.

THE LIGHT-ADAPTED EYE

In the light-adapted eye a more complicated set of mechanisms is necessary because as well as being a receiver of light the eye is able to distinguish between colours and also between the shape of one object and another. Thus the number of mechanisms used

The absorption spectrum of visual purple which may be extracted from the retinae of animals with digitonin solutions is shown in Fig. 2. The full line curve gives the absorption spectrum of the unbleached solution and the maximum in the visible at 500 mμ is due to the light-sensitive portion of the molecule.

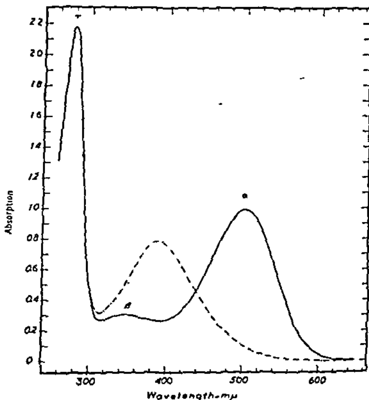


FIG. 2. (After Wald, 1949.) The absorption spectrum of a visual purple solution. Full line before exposure to light. Dotted line after exposure to light

The maximum at 280 mμ is due to the protein portion of the molecule and does not alter its position on illumination. The dotted curve shows the absorption of the pigment when it has been illuminated.

Suppose now that the sensitivity of the dark-adapted eye is the result of bleaching the single pigment visual purple; one would expect, *since light can only be effective if it is absorbed,*

that the spectral sensitivity curve of the dark-adapted eye would be similar in shape to the absorption spectrum of visual purple over the visible region of the spectrum, because at those wavelengths at which absorption is greatest there should clearly be the greatest sensitivity. Comparisons of this kind have been made a number of times during the last 100 years and one such is shown in Fig. 3.

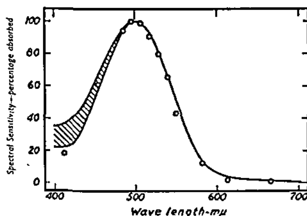


FIG. 3. (After Wald, 1938.) A comparison between the spectral sensitivity of the dark-adapted eye (points) and the absorption spectrum of visual purple (full line and hatched area) The sets of data have been equated at the maximum

In this diagram the points are the spectral sensitivity curve of the dark-adapted eye and the hatched area the limits within which the absorption spectrum of visual purple lies. The agreement between these two sets of data is so satisfactory that one can say that the dark-adapted eye uses a single mechanism and that that mechanism relies on the single pigment visual purple. The agreement between these two sets of data can be made even more perfect by making several minor corrections to the data.

THE LIGHT-ADAPTED EYE

In the light-adapted eye a *more complicated set of mechanisms* is necessary because as well as being a receiver of light the eye is able to distinguish between colours and also between the shape of one object and another. Thus the number of mechanisms used

must be more than one, since in all tasks requiring discrimination, one mechanism must behave in one way whilst others behave differently. If there are several mechanisms active no simple correspondence between the spectral sensitivity curve and the absorption spectrum of photopic pigments can be expected, but the sensitivity curves of each mechanism individually might sum to give the overall spectral sensitivity curve of the eye as given in Fig. 1 at (b). Since the visual pathway is anatomically divided into a number of stages: photochemical, neurone transmission, geniculate bodies and brain, one can look at the mechanism at each of these points to see what evidence there is which might point to an answer to the question: 'How many mechanisms are active?'

(a) *Histology of the Retina.* One might expect that the number of mechanisms active in the light-adapted eye would be demonstrable by histological methods; in fact one finds in retinal sections only two kinds of photoreceptors, rods and cones. The rods are the photoreceptors for night vision and are associated with visual purple and the cones in man are of one type only. In the fovea centralis, which is the region of the retina around the fixation point, and which has high visual acuity, the cones are the only type of receptors and although colours can be easily distinguished by the fovea only one type of receptor can be demonstrated. In the bipolar and ganglion cell layers differentiation into several series of different pathways is not apparent, although it is possible that the mop bipolar cell is associated with the rod pathway only. The reason for the lack of knowledge on the number of mechanisms in the anatomy of the retina is that the only adequate method of staining is the silver impregnation method, and this technique demonstrates the cells of the retina in a haphazard manner and does not show the way in which they are connected together. Indeed it is probably impossible to demonstrate these connexions histologically since it is the functioning of the synaptic tissue between the cells which divides the neural pathways into functional groups.

(b) *The Photochemical System.* Apart from visual purple, which has already been discussed, visual violet, a pigment similar to visual purple, and having similar functions in the dark-adapted

eyes of freshwater fish, the information on photosensitive photopic pigments is scanty. Wald (1949) has described a photosensitive pigment obtained from bleaching extracts of chicken retinae with red light which he called iodopsin; this pigment has an absorption maximum at about 565 mμ: Dartnall (1952), by a new method of differential bleaching, has described a further pigment with a maximum at 467 mμ. Dartnall's method is summarized in Fig. 4. He extracted the photosensitive material from whole tench retinae and exposed the solution to light of wavelength 610 mμ. This resulted in a partial bleaching of the

red-sensitive component of tench. They were found to be identical.

Having fully bleached a portion of the tench extract with 610 mμ he then proceeded to bleach the extract further with white light. Dartnall then obtained a further difference spectrum representing the photosensitivity of the residual material (the 'Red-insensitive' curve in Fig. 4). The sum of the red-insensitive and red-sensitive curves gives the open circles in Fig 4 and these figures show agreement with the closed circles and full line marked 'total' in Fig 4, which is the difference spectrum for the whole solution when bleached with white light. Dartnall thus effected a partition of the total photosensitivity of the tench extract into a component indistinguishable from visual violet and a new component having a maximum in the blue

authenticated visual pigments; visual purple and visual violet, which appear to be alternative pigments and associated with the rod receptors, and iodopsin and visual pigment 467. Only two of these pigments could be utilized by a photopic visual mechanism, and the evidence from the photochemical field would seem to indicate that two mechanisms are available for colour dis-

crimination, an inadequate number if other visual data are considered. This field of research is at the moment exceptionally active and there is little doubt that this conclusion will soon be

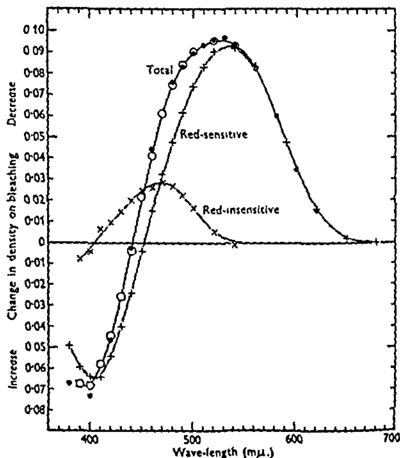


FIG 4 (After Dartnall, 1952) The results of an experiment on a solution from tench retinae showing the difference spectrum of the new photosensitive material, visual pigment 467, which is labelled in the diagram as the red-insensitive component

altered. Whether it is possible to have three or more photosensitive pigments in a single eye has yet to be discovered and the results can be awaited with interest.

(*c*) *The Nervous Transmission.* The most complete work on the

organization of the nervous mechanism of the eye in relation to the photopic mechanism has been done by Granit (1947). He devised a small platinum wire electrode which could be applied to the surface of the retina through the open bulb and could pick up the spike action potentials from single ganglion cells. These single fibre records could be displayed on the cathode ray

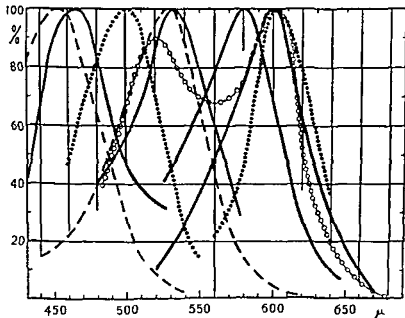


Fig. 5 (After Granit, 1947). The spectral sensitivity curves of modulators obtained by Granit by electro-physiological methods from a number of animals

tube. On stimulating the whole of the retinal surface with light of various wavelengths Granit was able to obtain sensitivity curves from the elements which he had picked up. He demon- [illegible]

lators'. Some of these curves are shown in Fig. 5. These narrow curves have been obtained by a number of different methods; by direct stimulation, light adaptation and polarization of the

retina with a steady current. Their maxima are usually located in three main retinal regions, orange, green and blue.

There has been considerable discussion as to whether Granit's results indicate the presence of three or more mechanisms in the photopic eye. Some would say that the three retinal regions correspond to a triple subdivision of the photopic eye and that animals would have perhaps one modulator from each of three regions. On the other hand it might be that an animal *could* have all the modulator systems found by Granit, in which case as many as eight seem to be possible, so that the information obtained from Granit's electro-physiological work does not answer the question, 'How many mechanisms?' in a simple way.

(*d*) *The Whole Visual Mechanism*. Subjective experiments on the human eye using the whole visual mechanism and the visual sensations might also be expected to reveal the number of mechanisms operative in the light-adapted state. Ever since the days of Helmholtz, who developed an idea originated by Thomas Young, the results of the colour-matching experiment have been taken as the basis for determining the mechanisms active in the light-adapted eye. In this experiment the subject is shown two adjacent light fields: one illuminated by suitably coloured light, the other by a mixture of light from three suitably chosen spectral regions, say a red, green and blue, the wavelengths of which remain standard throughout the experiment. The observer is asked to match in colour and brightness the two halves of the field by varying the amounts of red, green and blue light in the mixed field. A perfect match can be obtained for the great majority of colours, and since this is possible by the use of *three* lights, this is taken to indicate that there are three mechanisms active in the light-adapted eye. Recently the development of methods of measurement with a restricted visual field (15 minutes in diameter) has added some interesting information. If a colour-matching experiment is performed with a bipartite field subtending 15 minutes at the eye, the eye behaves as though it were partially colour-blind. This partial colour-blindness, known as tritanopia, was rediscovered by Willmer (1944) at the centre of the central fovea. Under these conditions of observation all the wavelengths of the spectrum can be matched with mixtures

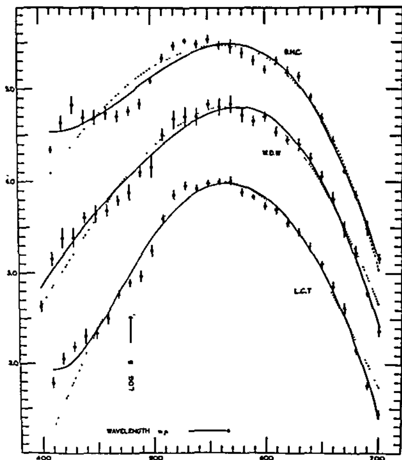


FIG. 6. (After Thomson, [illegible].) The spectral [illegible]

of two standard spectral lights, a red and a blue, and thus vision under these conditions is said to be dichromatic. Classically this state of affairs would indicate that two visual mechanisms are operative in the light-adapted central fovea.

When the spectral sensitivity of the foveal centre is measured with a similar small field, a series of inflections is found in the

curve, two of which have been known for some time. In Fig. 6 the logarithm of the spectral sensitivity is given on the ordinate and the points refer to three observers with normal colour vision The full lines indicate mathematically smooth curves with which the spectral sensitivity points themselves can be compared. Dips in the curve can be seen around wavelengths 600 mμ, 540 mμ, 480 mμ and 430 mμ.

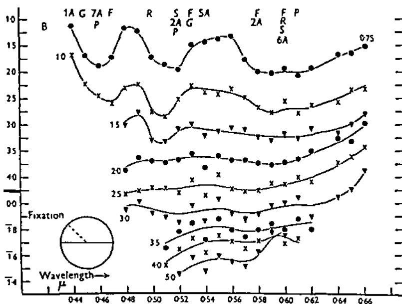
Since, however, the energy determinations for the light used in the measurement are made at the cornea, these sensitivity curves refer to the whole eye and no doubt some correction should be applied to them to allow for the light which is absorbed in the eye media themselves. It is possible that the inflections might disappear if this correction were known. However, so far as the information of absorption of light in pre-retinal media goes, there is insufficient irregularity to account for the shoulders seen in the measurements. If the irregularities are due to the imperfect fusion of the separate sensitivity curves of the colour mechanisms, then there would appear to be more than two of these active.

A further paradox appears when another experiment with the 15-minute field is considered. Fig. 7 shows the results of an experiment in which the ability to measure small differences of luminance (brightness) is measured with a 15-minute field. The ability to discriminate between one luminance placed on the lower half of the field and another slightly greater one placed on

sents a different overall brightness at which the measurements

instrument The curves have been artificially displaced, one above the other, to enable their detail to be seen. At high brightness the wavelength of the light makes little difference to the ability to distinguish between closely related intensities, but at low brightness the ability to distinguish luminance differences is best at wavelengths 460 mμ, 520 mμ and 600 mμ. These wave-

lengths correspond to those at which Granit has found the maximum sensitivity of colour modulators, and it would seem that the luminance discrimination is here the result of *three* separate mechanisms. This would be in direct conflict with the evidence



[illegible]

from the colour-matching experiments which indicate that with a field of 15 minutes in extent two mechanisms are at work. Contradictions of this kind as to number of mechanisms active in the light-adapted eye are a commonplace.

THE TRICHROMACY OF VISION

Even though there are a number of contradictions in the evidence of visual theory, the results of the colour-matching experiment mentioned earlier are outstanding in their importance. The fact that one can match all colours by manipulations involving only three standard lights shows that the visual path-

curve, two of which have been known for some time. In Fig. 6 the logarithm of the spectral sensitivity is given on the ordinate and the points refer to three observers with normal colour vision. The full lines indicate mathematically smooth curves with which the spectral sensitivity points themselves can be compared. Dips in the curve can be seen around wavelengths 600 mμ, 540 mμ, 480 mμ and 430 mμ.

Since, however, the energy determinations for the light used in the measurement are made at the cornea, these sensitivity curves refer to the whole eye and no doubt some correction should be applied to them to allow for the light which is absorbed in the eye media themselves. It is possible that the inflections might disappear if this correction were known. However, so far as the information of absorption of light in pre-retinal media goes, there is insufficient irregularity to account for the shoulders seen in the measurements. If the irregularities are due to the imperfect fusion of the separate sensitivity curves of the colour mechanisms, then there would appear to be more than two of these active.

A further paradox appears when another experiment with the 15-*minute field is considered. Fig. 7 shows the results of an* experiment in which the ability to measure small differences of luminance (brightness) is measured with a 15-minute field. The ability to discriminate between one luminance placed on the lower half of the field and another slightly greater one placed on [illegible]

sents a different overall brightness at which the measurements were made; for the upper curve the field as a whole was only just distinguishable from darkness itself, and in the lower curves the overall brightness was as great as could be obtained with the instrument. The curves have been artificially displaced, one above the other, to enable their detail to be seen. At high brightness the wavelength of the light makes little difference to the ability to distinguish between closely related intensities, but at low brightness the ability to distinguish luminance differences is best at wavelengths 460 mμ, 520 mμ and 600 mμ. These wave-

REFERENCES

BLISS, A. F. (1948). *J. biol Chem.* **176**, 563.
DARTNALL, H. J. A (1952) *J Physiol.* **116**, 257.
GRANIT, R. (1947) *Sensory Mechanisms of the Retina.* Oxford University Press, London.
THOMSON, L. C. (1949). *J. Physiol* **108**, 78.
— (1951). *J. Physiol.* **112**, 114.
WALD, G. (1938). *J. gen. Physiol.* **21**, 795
— (1949). *Docum. Ophthal.* **3**, 94.
WILLMER, E. N (1944). *Nature, Lond.* **153**, 774.
WRIGHT, W. D. (1946). *Researches in Normal and Defective Colour Vision* Kimpton, London.

way is somewhere or another limited to three components. Those who feel that vision can be explained best by postulating more than three mechanisms in the visual pathway have often disregarded, or failed to understand, the significance of the colour-matching information. The experiment tells us that the colour-matching procedure has only three independent variables, and it is thus certain that for this test the eye operates with only three mechanisms. The point is not easily understood because it is mathematical in nature. The restriction to three mechanisms is similar to the restrictions which a body has when it moves about in space. The body can move sideways, backwards and forwards or up and down, and any complicated movement can be resolved into differing amounts of these three. Thus the independent movements which the body can perform in space are restricted to three. The independent procedures in the colour-matching experiment are restricted to three, and this tells us for certain that the mechanisms somewhere in the visual pathway are limited to three.

Unfortunately, three mechanisms, which pass right through the visual system from the photoreceptors to the brain and remain discrete and independent throughout, cannot explain a large number of experimental facts, and we are forced to the conclusion that the only theories of vision which will turn out to be satisfactory are those which admit that there is a restriction to three mechanisms somewhere in the visual pathway, and that elsewhere the systems may be three or more. Theories of this kind are known as zone theories and it is undoubtedly in this direction that progress will in future be made.

investigation of the tissue changes which are induced by the injection of turpentine subcutaneously. Both demonstrate the early exudation of fluid from blood vessels in the affected region

TABLE 2. Early Inflammatory Reactions in Subcutaneous Tissue of Rats after Injection of Turpentine
(Compiled from Ernst, 1926)

5 minutes	Blood vessels slightly dilated, leucocytes increased within blood vessels
15 ,,	Blood vessels dilated, endothelial lining cells swollen, leucocytes definitely increased within vessels, oedema around vessels.
30 ,,	Pronounced vasodilatation, margination of leucocytes [illegible]
1 hour	[illegible] of [illegible]
2 hours	Much exudation of fluid, fibrin forming, marked migration of leucocytes.

followed later by the migration of leucocytes from these vessels. A third feature must be added, swelling of the endothelial cells which line the blood vessels, which is remarkably constant.

METHODS OF INVESTIGATION

These fundamental observations have come from a long line of studies which have resulted from classical methods of investigation which, I need scarcely remind you, depend upon the production of changes in living animals and the sacrificing of these animals at given intervals for the microscopical study in fixed preparations of what was happening at the time of death Such a static method has serious limitations which need not be entered into here. It is therefore all the more satisfactory that direct observations on living animals have brought out the same features attributable to injury which we had learnt from static investigations. Of these methods, I need only mention three.

The first may be dismissed in a few words though its contributions to our subject have been great. The idea of examining transparent living tissues by means of a microscope seems to have occurred first to Albrecht von Haller and Spallanzani towards the end of the eighteenth century though Malpighi had

VIII

Tissue Responses to Injury

G. R. CAMERON

We know quite well the microscopical picture of an abscess, a diphtheritic membrane or the early stages of a lobar pneumonia and are aware that each of these conditions shares two histological departures from the normal: (1) the presence of an unusually large number of leucocytes within the affected tissue, and (2) the presence of an exudate of plasma or modified plasma. So constant are these findings that we make use of their occurrence in a tissue in determining the presence of harmful agents, even when we know for sure that it is impossible to demonstrate such agents in the affected region. Experience, too, has taught us that a time-table may be constructed in which the leucocytes and fluid can be arranged in order of their appearance on the scene of damage; in other words, the time sequence of the phenomena of inflammation can be established with accuracy. Two such time-tables are illustrated here. The first one is based on the careful study by Carscadden (1927) of what happens when the liver is incised aseptically, the second table is compiled from Ernst's (1926)

TABLE 1 Early Inflammatory Reactions in the Liver of Rabbits, after Aseptic Incisions
(Compiled from Carscadden, 1927)

3 minutes	Liver cells injured, exudation of plasma, fibrin forming.
30 ,,	Margination and stasis in blood vessels. Much fibrin Dilated vessels
1 hour	Much leucocytic migration Liver cells disintegrating.
2 hours	Masses of leucocytes in injured tissue
3 ,,	Much exudate and fibrin, clumps of leucocytes

to leave the capillaries of an inflamed area within a few minutes of injury, and goes on exuding for a long time, sometimes for days or weeks on end. Studies with large molecular dyes which attach themselves to plasma proteins suggest that the permeability of these capillaries is rapidly increased in the neighbourhood of an injury and that this permeability alteration persists even when the flow of blood through vessels is slowed down or ceases. There is no denying that hydrostatic factors may influence the exudation of fluid but they are not the major factors. So, too, the migration of leucocytes from vessels is not merely a mechanical process as some writers would have us believe though it may be modified by such factors. Direct observation shows quite plainly that leucocytes can move in a fluid medium independently of the rules of flow of the fluid, proceeding by a minimal exercise of energy in a directed or polarized fashion towards the site of tissue damage. Opinions have been expressed from time to time about the meaning of these phenomena. Exudation of fluid has been suggested as a means of flushing the affected site of injurious agencies and products, of bringing to bear upon bacteria antibodies of all kinds and of leading to the temporary blockade of the region by the formation of a solid phase, fibrin, as the result of coagulation of one of the components of the exuded fluid. Thus lymphatics may be blocked up with this fibrin mass so that lymph flow stops and one of the exits from the inflamed region thus may be closed. All such developments have been held by first one and another worker to be of advantage to the organism because of the sealing off of the danger spot in some such fashion.

MIGRATION OF LEUCOCYTES

Even greater flights of imagination have been associated with the migration of leucocytes. Through the life work of the great Russian biologist, Metchnikoff, we have been convinced of the universal occurrence of phagocytosis in living things. I need only recall the behaviour of unicellular organisms in their blind search for food, the apparently purposive response of the lowly slime moulds to things which harm them, the setting aside of motile cells in the coelomic spaces of coelenterates which hurry

long before used the method in his wonderful demonstration of capillaries. A series of studies on the frog's mesentery or web carried out in Great Britain by John Thomson of Edinburgh (1813), Charles Hastings (1820), William Addison (1843), Thomas Wharton Jones (1850) and Lord Lister (1858), followed by the more penetrating investigations of Julius Cohnheim (1867) in Germany, taught us to expect vascular phenomena, exudation of fluid and migration of leucocytes from vessels in the vicinity of cell injury, and from these observations came the explanation of the time-honoured naked eye signs of inflammation.

In recent years, it has become possible to watch what goes on in mammalian tissues, e.g. the rabbit's ear, by means of the ingenious Clark-Sandison chamber. My assistant, Dr. A. D. Bangham, has modified this chamber by the insertion of an electrical filament so that standard burns may be produced in the tissues under observation at will and the effects of the burn then followed as long as one wishes in unanaesthetized rabbits. [illegible] sequence [illegible] that pre- [illegible] Oxford, present an unforgettable permanent record of reactions to injury.

The third method, introduced by Speidel in America, allows of the cinematography of inflammation in the transparent tail of a tadpole for weeks or months on end. I must refer the reader to the original papers of these workers for details and conclude this section by stating that all such studies have given ample confirmation of the observations and conclusions made by the early investigators. Thus there is overwhelming proof of the phenomena of inflammation and there is no gainsaying the time-tables which have been compiled by many workers. The first steps in the analysis of tissue response to injury were taken in this fashion.

EXUDATION OF FLUID

But analysis leads to further analysis and it is not surprising that exudation of fluid and migration have in turn posed many questions. A great deal of careful work has shown that fluid begins

to leave the capillaries of an inflamed area within a few minutes of injury, and goes on exuding for a long time, sometimes for days or weeks on end. Studies with large molecular dyes which attach themselves to plasma proteins suggest that the permeability of these capillaries is rapidly increased in the neighbourhood of an injury and that this permeability alteration persists even when the flow of blood through vessels is slowed down or ceases. There is no denying that hydrostatic factors may influence the exudation of fluid but they are not the major factors. So, too, the migration of leucocytes from vessels is not merely a mechanical process as some writers would have us believe though it may be modified by such factors. Direct observation shows quite plainly that leucocytes can move in a fluid medium independently of the rules of flow of the fluid, proceeding by a minimal exercise of energy in a directed or polarized fashion towards the site of tissue damage. Opinions have been expressed from time to time about the meaning of these phenomena Exudation of fluid has been suggested as a means of flushing the affected site of injurious agencies and products, of bringing to bear upon bacteria antibodies of all kinds and of leading to the temporary blockade of the region by the formation of a solid phase, fibrin, as the result of coagulation of one of the components of the exuded fluid. Thus lymphatics may be blocked up with this fibrin mass so that lymph flow stops and one of the exits from the inflamed region thus may be closed. All such developments have been held by first one and another worker to be of advantage to the organism because of the sealing off of the danger spot in some such fashion.

MIGRATION OF LEUCOCYTES

Even greater flights of imagination have been associated with the migration of leucocytes. Through the life work of the great Russian biologist, Metchnikoff, we have been convinced of the universal occurrence of phagocytosis in living things. I need only recall the behaviour of unicellular organisms in their blind search for food, the apparently purposive response of the lowly slime moulds to things which harm them, the setting aside of motile cells in the coelomic spaces of coelenterates which hurry

to the site of activity of injurious agents, to illustrate how early on in the evolutionary order this phenomenon appears. But we should not forget that phagocytosis is not the only means of protection which these creatures possess. Let me illustrate my warning by referring to injury response in earthworms and caterpillars.

If foreign material such as a suspension of India ink or a bacterial culture be injected with care into the coelomic cavity of a segment of earthworm several responses are evoked almost at once. A certain amount of the suspension is squirted out of the coelom by way of the dorsal pore of the segment and is lost. Some is passed on to the distal segment through the paired openings which join the segments until it reaches the tail segment. There it accumulates as clumps which become surrounded by phagocytes until of macroscopical size, the so-called 'brown bodies'. Eventually it is got rid of by the device of autotomy, the process of snapping off terminal segments and their replacement by new formation of segments. A little is also expelled through the paired nephridia of each segment. Finally, quite an amount seems to precipitate and clump within the cavity of the injected segment close to the outer wall of the segment in contact with

and sets free hordes of coelomic phagocytes and forms large plates of these cells in close relationship to the clumped ink or bacteria. Within a short period of time much of the foreign material is taken up and enclosed in a dense mass of phagocytes. We have here a process analogous to the formation of an abscess or a tubercle (Cameron,
sists for a long time, b
wandering away of pha
material.

Response to injected material, whether inert or living, is likewise a complicated affair in insect larvae. That pest of beehives, the wax moth caterpillar (*Galleria mellonella*), is well suited for such studies (Metalnikov, 1927; Cameron, 1934). Let me remind the reader, first, of its simple anatomy. In the fully-grown state it may be nearly 2 cm in length and 0 5 cm broad,

and it possesses a thin chitinous wall through which can be inserted a glass capillary tube without much difficulty. In this way aqueous suspensions of bacteria, dyes, particulates or cells can be injected in amounts which will not harm the caterpillar. Such a suspension is rapidly dispersed through the two main compartments because of the constant circulation of the coelomic fluid maintained by the simple contraction of the rather primitive heart. This structure lies within the posterior coelomic compartment suspended upon a broad fenestrated membrane called the pericardium. Coelomic fluid has to pass through the tiny openings of the pericardium on its way to and from the heart and in so doing it glides along the surface of large phagocytic pericardial cells. These may be considered as a primitive reticulo-endothelial system for they clear the coelomic fluid of much that is abnormal in its physical composition. Injected carmine particles or bacteria are stored in great numbers in the pericardial cells and they may persist here for the remainder of the caterpillar's life, even surviving the vicissitudes of insect metamorphosis. I have demonstrated, for instance, living tubercle bacilli in the pericardial cells right through the pupal stage to the emergence of the moth after such bacilli are injected into the coelom of the larva. The coelomic fluid also contains many phagocytic cells which can take up particles at a great rate. The Malpighian tubules, which are primitive kidneys, also play a part in defence by excreting foreign material and finally an encapsulation process, like that I have described in the earthworm, accounts for much of the introduced material These studies again show that phagocytosis is one of several mechanisms invented by the living organism in reply to injury. Phagocytosis varies greatly in its efficiency in these simple forms and one gains the impression that it is a sluggish business which is not especially effective.

I need not multiply my examples of this type of response and the deductions which have come from them, inspired by the enthusiasm of Metchnikoff, but I must insist that it is only too easy to read into the fascinating phenomenon of phagocytosis a great deal more than is justified. In man there is no doubt whatsoever that it is a common, everyday process which may be of

great value in dealing with relatively small numbers of bacteria and foreign material, but one must still doubt its efficiency in the presence of a virulent micro-organism, a heavy sowing of germs, or a particularly severe or extensive destruction of tissue.

From all such researches and discussions have come several general rules and simplifications which may be stated as follows:

(1) Agents which kill tissues produce inflammation.

(2) Agents which induce inflammation can produce damage or death of tissue cells.

(3) A great variety of agents produce the same phenomena of inflammation and the only factor common to all of them is that of injuring cells or cellular products.

(4) Inflammation goes on independently of any controlling action of the central nervous system although it may be modified by nervous factors.

(5) Diffusible substances can be demonstrated in extracts of dead tissue or in the products of inflamed tissue, which lead to the phenomena of inflammation when they are introduced into healthy living tissues in very small amounts.

DIFFUSIBLE SUBSTANCES

It is now a matter of history how Lewis and his school built up the argument, from ingenious researches on man, that a histamine-like substance or substances can account for the exudation of fluid and the vascular phenomena of acute inflammation. They obtained highly suggestive evidence that H substance is liberated from damaged cells in amounts sufficient to evoke the vascular responses of inflammation. The Lewis school also showed that histamine was not responsible for the migration of leucocytes (Grant and Jones, 1929) so that search for some other factor became necessary.

During 1936, Menkin (summary, 1940) demonstrated that cell-free inflammatory exudate from a number of sources possesses the property of increasing capillary permeability and of inducing polymorphon[illegible] migration from skin blood vessels in the rabbi[illegible] blood albumin has similar [illegible] and Cullumbine and Rydon (1946) extended the latter observa-

tions to several enzymes and proteins and made partially successful efforts to purify the active principle. They concluded, like Menkin, that this was a polypeptide and accepted his name 'leucotaxine'. None of these workers gave any evidence as to whether these biological properties were confined to one compound or, as seemed more likely, were shared by many.

Recently my colleague Dr. W. G. Spector (1951) has re-investigated the properties of certain enzymic digests of protein which reproduce the crucial features of acute inflammation and has shown that the active principle of all such digests is the peptide fraction. Working with a peptic digest of fibrin, Spector found that the properties of increasing capillary permeability, causing vascular endothelium to swell and inducing leucocytic migration from the blood vessels, are shared by peptides with an average chain length of 8–14 amino-acid residues. From the same digest a fraction was obtained, with an average peptide chain of 5 amino-acids, which could induce polymorph emigration and endothelial swelling but had no action on vessel permeability. A pure crystalline peptide, pancreatic trypsin inhibitor, with an estimated molecular weight of 6000, was shown to possess all three of the biological properties. Spector concludes that such properties are possessed not by one but by many peptides within certain limits of molecular size and that the biological action may vary with the length of the peptide chain. He has succeeded in demonstrating in an inflammatory exudate obtained from a goat that the active principle of all three properties displayed by the exudate is the peptide fraction. With proper caution Spector maintains that higher peptides are concerned in producing some of the changes of acute inflammation, especially the perpetuation of leucocytic migration, but he rightly points out that increased capillary permeability and the consequent exudation of fluid which set in very rapidly after injury are quite likely the direct result of the trauma or infection. Here liberation of histamine may be the responsible mechanism or a physical alteration in the protoplasm of damaged cells may suffice.

We know very little, too, about the factors which are responsible for the liberation of histamine and peptides from injured

cells or their products. Several pointers to a profitable field of investigation come to mind.

(1) It is well known that certain bacteria of the Clostridium group, e.g. *Cl. welchii* type A, produce powerful lecithinases (α-toxin) and collagenases capable of hydrolysing compounds which are essential constituents of all cell membranes. Thus they may lead to disruption of muscle cells and destruction of reticulum and collagen fibres (see the excellent review of Payling Wright, 1950).

(2) Some bacterial toxins, as well as snake and bee venoms, liberate histamine from organs perfused with them (Feldberg and Kellaway, 1937, 1938). This effect is due in some cases to a lecithin-destroying agent, lysocithin.

(3) Rocha e Silva (1940) has shown that trypsin can bring about histamine liberation *in vivo* and this suggests that perhaps the activation of proteolytic cell enzymes may in similar fashion set free *H* substances.

(4) Beloff and Peters (1945, 1946) have identified proteases in burned skin which act on proteins to produce active peptides (Cullumbine *et al.*, 1947).

(5) The recent studies of Feldberg and Paton (1951) on histamine liberators suggest that histamine may be set free from cells in certain parts of the body without any gross structural disturbance of the cell. This discovery will be followed with great interest by pathologists who have been puzzled from time to time by the progress of an inflammatory reaction in the apparent absence of tissue injury.

From these somewhat disconnected facts has been concocted a story something like the following. Injury of various types so acts upon tissue cells or their products that histamine, peptides and perhaps other diffusible compounds are rapidly set free and diffuse into the tissue spaces. These modify the permeability of the adjacent capillaries and small vessels and dilate arterioles whereby blood flow is at first accelerated and then slowed, plasma exudes from the capillaries and somewhat later leucocytes migrate towards the site of injury. The exudation of fluid dilutes the injurious agent and the products of its activity, washes away these harmful substances, leads to the production

of a fibrinous gel which fills up many of the tissue crevasses and lymphatics and so prevents the dispersion of harmful material, and brings into play immunity bodies which may have been developed against agents such as bacteria. Migration of cells is also a protective affair since it results in the mobilization of large numbers of phagocytes at the spot where they are most needed.

IS INFLAMMATION BENEFICIAL?

But is this story true? What reasons have we for regarding inflammation as a purposive phenomenon? Not very many which are convincing, I fear. Let me consider some of these in turn, together with their counter-arguments.

(1) Exudation of fluid is said to dilute harmful material, especially when there is ample space for dilution or provided means exist for the removal of the diluting fluid. But the exudate may sweep away the injurious material to a wider field of activity. This surely is highly undesirable in the case of virulent bacteria, even though further protective devices, such as the clearing mechanisms of the reticulo-endothelial system, have been invented by the organism.

(2) It is said that the formation of fibrin by clotting of the exudate is of benefit through the blockade of pathways leading from the inflamed area. Fibrin, however, is a good culture medium for many bacteria and may favour their growth locally. Many observers, too, have not been impressed by morphological evidence of the blockade phenomenon. In any case it is slow in developing and often is delayed until too late to hold up the spread of bacteria.

(3) Much emphasis has been placed upon the presence of fibrin thrombi within lymphatics of an inflamed area. Such thrombi are said to occlude lymphatics and suppress the flow of lymph from the area But direct measurements of lymph flow from a severely injured region have shown that there is increased lymph flow for quite a long time (Drinker and Field, 1933; Cameron *et al.*, 1947), so that this suggestion has no sound foundation.

(4) Fibrin is claimed to favour the activity of phagocytes through acting as a bridge along which phagocytes may move

towards their prey. There is some evidence for this view, but the mechanism applies to the later stages only, for fibrin is formed slowly. At the very moment when protection against the spread of the injurious agent is most needed, this particular device is not available.

(5) An ingenious suggestion attributes to the protein of the exudate the power of precipitating bacteria and other harmful matter just as salt and dyes are precipitated in the presence of protein. May not this also favour localization of bacteria at the site of activity or concentration there of toxic substances?

(6) A serious criticism of the theory is that no convincing quantitative proof exists of the value of phagocytosis in living tissue, especially against large numbers of bacteria, or against virulent strains of bacteria.

(7) Much striking experimental evidence has been collected which shows that during the early phase of inflammation the affected tissue is no more resistant to the absorption of bacteria than normal tissue. The time required for an effective amount of inflammation to develop about invasive bacteria is much longer than the time required for bacteria to spread from the site at which they lodge (Issaeff, 1894; Cobbett and Melsome, 1898; Wadsworth, 1904; Rivers and Tillett, 1925; Opie, 1929, 1930; Cannon and Pacheco, 1930; Angevine, 1934, 1936; Rich, 1936; Lurie, 1936; Cannon and Hartley, 1938).

Only when the exudate contains specific antibodies against the invaders do we find convincing evidence for the view that inflammation is protective in function. I would therefore suggest that inflammation is of value when it leads to the rapid accumulation of antibodies at a spot where they are urgently needed; under other circumstances it may be a nuisance or even a menace. It is a kind of evolutionary gloss, a blundering, clumsy affair which can add to the burden of the already over-taxed organism but which may turn out to be an advantage if the organism has already learnt the trick of providing the most successful of all safeguards, immune bodies.

Nearly 150 years ago, the great French experimental physiologist Magendie, then a young man of twenty-six, wrote his first important scientific paper. In it he pointed out that it was the mark of a perfected science to have a small number of prin-

ciples under which a large number of facts might be readily grouped. In that respect the physical sciences—chemistry, physics and astronomy—and above all the queen of sciences, mathematics, are models of what science ought to be. How often must the thoughtful student of disease have asked himself whether pathology is such a science? Can many of the facts, accumulated so patiently over the years, be grouped together under a few principles? Are we in possession of principles which allow of the neat, concise summary of those features which we hold to be significant and from which we may venture into the field of deduction and speculation? The purpose of this lecture has been to show that in one field of pathology we may claim modest success in this aim, even though the pattern is still indistinct in many of its details.

REFERENCES

ADDISON, W (1843). *Experimental and Practical Researches on Inflammation*, etc. London.

ANGEVINE, D M (1934), *J. exp. Med.* **60**, 269

— (1936) *J. exp. Med.* **64**, 131.

BELOFF, A, and PETERS, R A (1945) *J. Physiol* **103**, 461

— (1946) *J Physiol.* **105**, 54

CAMERON, G. R (1932) *J Path. Bact* **35**, 933

— (1934) *J Path Bact* **38**, 441

— COURTICE, F. C, and SHORT, R. H. D (1947) *Quart J exp. Physiol.* **34**, 1

CANNON, P. R, and HARTLEY, G. (1938). *Amer J. Path.* **14**, 87

— and PACHECO, G A (1930). *Amer J. Path* **6**, 749

CARSCADDEN, W. G. (1927). *Arch Path* **4**, 329

COBBETT, L, and MELSOME, W. S (1898). *Zbl allg Path path. Anat* **9**, 827

COHNHEIM, J (1867). *Arch path. Anat* **40**, 1.

CULLUMBINE, H, McDONALD, F, and SIMPSON, M. M (1947). *J. Path Bact.* **59**, 467

— and RYDON, H N (1946) *Brit J. exp. Path.* **27**, 33.

DRINKER, C K, and FIELD, M E (1933) *Lymphatics, lymph and tissue fluid*. London

DUTHIE, E S, and Chain, E (1939). *Brit J exp Path.* **20**, 417

ERNST, T (1926) *Beitr path Anat* **75**, 229

Feldberg, W. (1938). *Aust. J. exp. Biol. med. Sci.* **16**, 249.
— and [illegible]

[illegible]

etc. London.
Issaeff, (1894). *Ztschr. Hyg.* **16**, 287.
Jones, T. Wharton (1850). *Guy's Hosp Rep* **7**, 1.
Lister, Lord (1858). *Phil. Trans.* **148**, 678
Lurie, M B (1936) *J. exp Med.* **63**, 923
M [illegible]
M [illegible]

Opie, E. L. (1930). *The Harvey Lectures*, **24**, 197.
— (1929). *J. Immunol.* **17**, 329.
Rich, A. R (1936). *Arch. Path.* **22**, 228.
Rivers, T. M., and Tillett, W. S. (1925). *J. exp. Med* **41**, 185.
Rocha e Silva, M. (1940). *Arch. exp Path* **194**, 335, 351.
Spector, W. G. (1951). *J. Path. Bact.* **63**, 93
[illegible]

IX

Studies of Normal and Pathological Physiology of the Kidney

D. D. VAN SLYKE

It is a peculiar honor for the representative of a distant laboratory to be invited to discuss renal physiology in the land of Bowman, Bright, Cushny, Starling, and the successors who carry on their great tradition. The data that I shall present are in part the yield of some years of experimental and clinical observation by a group of investigators at the Hospital of the Rockefeller Institute. But the work cannot be offered without outlining the classic studies that provided its foundation, and including results of contemporary colleagues whose explorations either inspired our embarking on new ventures, or carried on where our navigation reached its limit.

THE FILTRATION-REABSORPTION PICTURE

Since the appearance of Cushny's *The Secretion of Urine* in 1917, the 'filtration-reabsorption' theory has provided the working hypothesis for a large part of the research done on the physiology of the kidney, and has formed a picture into which an otherwise disconnected mass of facts could be fitted in logical order Without such a picture it would be difficult to integrate the results that I shall try to review. I therefore pause to sketch its outlines as it stands today, with some of the added background and nuances that have grown into the picture since it was first drawn by Cushny's master hand

About 20 per cent of our bodies consists of extracellular fluid which surrounds and bathes the cells. This fluid is of extra-

ordinarily constant composition with regard to its concentration of crystalloid solutes. The organ by which this constant composition is chiefly maintained is the kidney. It maintains not only the composition, but also, by regulating the water output, the volume.

According to the filtration-reabsorption theory, as the blood courses through the glomeruli, at a rate in adult man of about a liter per minute, part of the plasma water (about one-fifth, as will be shown later) with its dissolved crystalloid solutes is filtered out and passes down the tubules. In the tubules there occurs a reabsorption of enough water to maintain the extracellular fluids at their normal volume, and enough of the various crystalloids to maintain their individual concentrations in the extracellular fluid constant within extraordinarily narrow limits.

The intelligence with which the tubular cells discriminate between the different constituents of the glomerular filtrate and decide which to reject and which to reabsorb is almost unbelievable. The tubules divide the constituents of the glomerular filtrate into three classes. First, those that the organism needs to retain as completely as possible, such as glucose and amino-acids, which are reabsorbed almost completely. Second, those constituents of the glomerular filtrate which must be reabsorbed in varying amounts to retain exactly enough to preserve the constant volume and electrolyte content of the extravascular body fluid. Such are the water and the electrolytes. Third, those metabolites for which the organism has no further use, such as urea, uric acid, and creatinine, which are rejected and excreted.

The cells of the tubular epithelium act like small force pumps; they pump water from the glomerular filtrate back into the blood against a head of osmotic pressure which becomes increasingly great as the filtrate goes down the tubule and becomes more concentrated. At the same time they selectively reabsorb or reject the different dissolved substances in the filtrate, according to whether the body requires these substances to maintain the composition and volume of the extravascular body of fluid. This selective pumping process is carried on until the urine is concentrated to an osmotic pressure which is usually many times greater than that of the blood plasma. For this

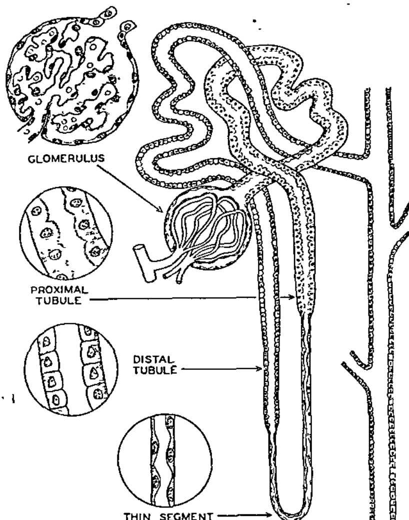


FIG. 1. Diagram of nephron. From Homer Smith (1951).

process against a head of osmotic pressure energy is required, as it is required to pump water uphill against a head of gravity. This energy is provided by the combustion of organic substances in the tubular epithelium. In fact, the rate of oxygen consumption of the kidney shows that it produces more than 100 times

ordinarily constant composition with regard to its concentration of crystalloid solutes. The organ by which this constant composition is chiefly maintained is the kidney. It maintains not only the composition, but also, by *regulating the water output*, the volume.

According to the filtration-reabsorption theory, as the blood courses through the glomeruli, at a rate in adult man of about a liter per minute, part of the plasma water (about one-fifth, as will be shown later) with its dissolved crystalloid solutes is filtered out and passes down the tubules. In the tubules there occurs a reabsorption of enough water to maintain the extracellular fluids at their normal volume, and enough of the various crystalloids to maintain their individual concentrations in the extracellular fluid constant within extraordinarily narrow limits.

The intelligence with which the tubular cells discriminate between the different constituents of the glomerular filtrate and decide which to reject and which to reabsorb is almost unbelievable. The tubules divide the constituents of the glomerular filtrate into three classes. First, those that the organism needs to retain as completely as possible, such as glucose and amino-acids, which are reabsorbed almost completely. Second, those constituents of the glomerular filtrate which must be reabsorbed in varying amounts to retain exactly enough to preserve the constant volume and electrolyte content of the extravascular body fluid. Such are the water and the electrolytes. Third, those metabolites for which the organism has no further use, such as urea, uric acid, and creatinine, which are rejected and excreted.

The cells of the tubular epithelium act like small force pumps; they pump water from the glomerular filtrate back into the blood against a head of osmotic pressure which becomes increasingly great as the filtrate goes down the tubule and becomes more concentrated. At the same time they selectively reabsorb or reject the different dissolved substances in the filtrate, according to whether the body requires these substances to maintain the composition and volume of the extravascular body of fluid This selective pumping process is carried on until the urine is concentrated to an osmotic pressure which is usually many times greater than that of the blood plasma. For this

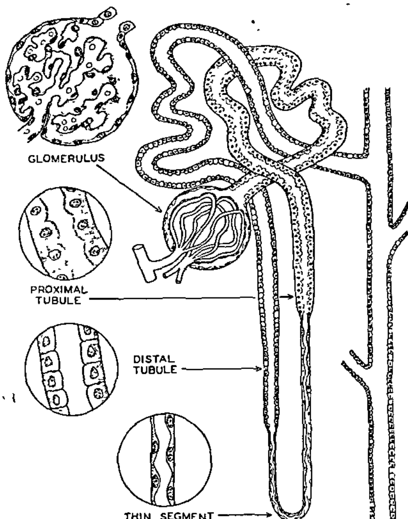


FIG 1. Diagram of nephron. From Homer Smith (1951).

process against a head of osmotic pressure energy is required, as it is required to pump water uphill against a head of gravity. This energy is provided by the combustion of organic substances in the tubular epithelium In fact, the rate of oxygen consumption of the kidney shows that it produces more than 100 times

the amount of energy that would be required to concentrate the urine if the tubular cells were functioning as perfect engines. The use that is made of the surplus energy is one of the problems of renal physiology that yet remains to be solved.

The renal tubules, in order to know to what extent they must selectively reabsorb the filtrate constituents of the second group, to maintain constant volume and composition of the milieu interieur, must receive messages from other parts of the body. These messages are at present not completely understood, but it is known that some of them are in the form of hormones. The pituitary secretes a hormone that stimulates reabsorption of water; when this hormone is lacking, diabetes insipidus results, with a flow of urine that may exceed 20 liters per day and result in dehydration of the body unless water intake keeps pace with such an output. Adrenal cortical hormones, at least partly controlled in their output by the pituitary, regulate the activity of the tubules with regard to reabsorption of sodium and potassium; in Addison's disease the deficit of messages to the tubules in the form of this hormone or hormones results in uncontrolled excretion of sodium and resultant sodium deficit in the extracellular fluid, which is responsible for part of the symptoms of the disease.

The filtration-reabsorption theory that I have sketched may be considered to have originated with the physiologist, Ludwig, a century ago. Ludwig (1844), from the structure of the nephron, proposed the theory that the glomeruli filter water and its diffusible solutes, and that the volume of the filtrate is decreased by reabsorption as the filtrate passes down the tubule. Cushny (1917), having more complete analyses of blood and urine at his disposal, added the conception that the reabsorption is not a simple osmotic diffusion, but a selective process, absorbing different proportions of different filtrate constituents according to the needs of the organism In opposition to this conception was that of Ludwig's contemporary, Heidenhain, who believed that the glomeruli filter only as much water as appears in the urine, and that the solid constituents are chiefly added to the filtrate by secretion from the tubular epithelium. These rival conceptions were reviewed by A. N. Richards (1938) in his

Croonian Lecture before the Royal College of Physicians. At the present day it may be said that observed facts support Cushny in so far as the excretion of water and the usual solutes of the urine is concerned, but that the tubules, besides their regularly exerted ability to reabsorb, appear to have the latent ability also to excrete when certain foreign substances, such as diodrast, para-amino hippurate, or penicillin are presented to the kidney. Then tubular excretion added to glomerular filtration provides an extraordinarily rapid removal of these substances from the body. When excessive amounts of potassium are administered, the tubules appear to be able to change from reabsorption to excretion of this element (Keith, King and Osterberg, 1943).

Confirmatory evidence for the 'modern' filtration-reabsorption theory of Cushny was afforded by the brilliant experiments of A. N. Richards and his school. Richards, with Wearn, Montgomery, Walker, and other collaborators (1938), developed a series of extraordinary methods for ultramicro analysis of the constituents of the blood plasma. These investigators also developed a technique whereby they could puncture a single glomerulus or different portions of the tubule of a nephron and withdraw the glomerular filtrate as it was formed. With their ultramicro methods of analysis they could determine the constituents of the glomerular filtrate in volumes as small as one cubic millimeter. They showed that chloride, bicarbonate, creatinine, urea, uric acid, pH, and glucose in the glomerular filtrate thus obtained had practically the same concentration as in the plasma. Thus, they demonstrated by direct analysis that the glomerular filtrate in truth is a simple filtrate of the plasma. The glomerulus functions as a mechanical filter, the blood pressure forcing a certain proportion of the water of the plasma, together with the crystalloid solutes that are dissolved in it, into the glomerular capsule by the hydraulic force of the blood pressure. Analyses of filtrate from the tubules showed that in the proximal tubules an isosmotic reabsorption of a large part of the filtrate and part of its solutes occurred; increase in osmolar concentration by selective reabsorption of water occurred in the loop of Henle or the distal tubule. These data were marshalled by Richards (1938) thirteen years ago in his Croonian Lecture,

and more have been added since, especially by Walker and Oliver (1941) who have extended the technique to mammalian nephrons.

One of the objections earlier raised to the filtration-reabsorption theory was that the theory requires the filtration and reabsorption of unbelievably great volumes of water, about 170 liters per day by the adult man. However, Walker, Bott, Oliver and MacDowell (1941), using Richards' procedure for direct catheterization of the tubules, were able to show in 1941 that in mammalian kidneys the amount of fluid passing from the glomerulus into the proximal tubule was equal to that required by the filtration theory, and that at least two-thirds of the water and all of the sugar were reabsorbed before the end of the proximal segment was reached.

THE PROPORTION OF PLASMA WATER FILTERED IN THE GLOMERULI

The group at the Rockefeller Hospital has been able to add data concerning some of the factors postulated by the filtration-reabsorption theory. One of these questions was the percentage of plasma water that is filtered into the glomerulus. This could

TABLE 1. Percentages of Ferrocyanide, Creatinine, and Inulin extracted from the Plasma of Blood flowing through Dogs' Kidneys

Excretory substance	Number of determinations	Mean percentage extracted	Standard deviation from mean
Ferrocyanide	25	18·8	5·5
Creatinine	36	19·9	3·8
Inulin	21	22·3	7·9

be determined by measuring the percentage of some plasma constituent that is filtered with the water, but not reabsorbed at all in the tubules Since 98 to 99 per cent of the filtered water is reabsorbed, the percentage fall in plasma concentration of such a substance occurring as the blood passes through the kid-

neys could indicate the percentage of the plasma water with which the substance was filtered. The anatomist, Gersh (1933–34), showed that for the rabbit such a substance is sodium ferrocyanide. During ferrocyanide excretion, sensitive histochemical methods showed that it was present in glomerular filtrate but was absent in the tubular cells, indicating that it was not passing through them in either direction.

In order to find out how much the concentration of ferrocyanide in the plasma decreased as the blood flowed through the kidney, it was necessary to obtain blood from the renal vein for analysis. Also, it was desirable that the experiments should be done under as nearly as possible normal conditions without operative disturbance and without anaesthesia. A procedure for accomplishing this was developed by Dr. C. P. Rhoads, now head of the Memorial Cancer Institute in New York Rhoads (1934) designed an operation whereby the kidney of a dog could be drawn out under the skin of the flank and so placed that the renal vein could be punctured by a needle by a technique similar to that by which blood is ordinarily drawn from a vein in the arm. The fact that ferrocyanide was handled in the kidneys of the dog, as Gersh had shown it to be handled in the kidneys of the rabbit without any reabsorption, was demonstrated by Dr. Benjamin Miller, then in our laboratory, who went to Gersh's laboratory and with Gersh performed experiments on the dog's kidney similar to those that had been done by Gersh with the rabbit. As shown in Table 1 (Van Slyke, Hiller and Miller, 1935b), the Rockefeller group found that a decrease of about 20 per cent occurred in the plasma ferrocyanide concentration during perfusion of the kidney, and it was deduced that about 20 per cent of the plasma water was filtered. Creatinine and inulin showed, within the limits of experimental variation, the same behavior as ferrocyanide. Richards, Bott and Westfall (1938) later provided independent evidence that inulin is excreted exclusively by glomerular filtration, thereby strengthening the deduction that had been based primarily on the behavior of ferrocyanide.

EXCRETION BY THE TUBULES

In mammals the processes of filtration by the glomeruli and selective reabsorption by the tubules appear to constitute the mechanism by which the composition of the body fluids is ordinarily maintained. However, there are aglomerular animals (Smith, 1951) in which all the excretory work is carried on by tubules alone, and it appears that the power of tubular excretion is retained by mammals, although used to a significant extent only to eliminate foreign substances, or to assist in elimination of normal excretory products under unusual conditions. One might consider tubular excretion as an hereditary reserve which can upon occasion be called into action. Thus, in the case of potassium there is ordinarily no evidence of tubular excretion; but when large amounts of potassium have been administered to a dehydrated subject (Keith, King and Osterberg, 1943) or in renal disease (Platt, 1950) the potassium clearance (*vide infra*) may be greater than can be accounted for by glomerular filtration alone, as measured by the inulin clearance, and the excess potassium excretion must be attributed to tubular excretion. In man creatinine shows a similar behavior (Miller and Winkler, 1937; Shannon, 1938).

The most marked examples of tubular excretion are shown in the elimination of foreign substances. One such substance is the dye, phenol red. Marshall (1931–32) showed that about 60 per cent of the phenol red in blood was removed as the blood flowed through the mammalian kidney. Such a complete removal could not be accomplished by filtration alone. In fact, very little of the dye exists in the free state in the plasma, most of it being bound to the proteins, so that little is presumably filtered. Other substances have been found which are removed so completely from the renal blood plasma that one must conclude that they also are removed by active secretion by the tubules. The proportion of hippurate extracted from the plasma by the kidneys was later measured in dogs by Phillips *et al.* (1946), and in men by Warren *et al.* (1944), by comparing the concentration of hippurate in the arterial plasma with the concentration in renal venous plasma, and was found in both dogs and men to average

normally about 87 per cent, when the arterial plasma concentration of the substance did not exceed 4 or 5 mg per 100 cc. Bradley (1948) finds by catheterizing the *right* renal vein of men an average extraction of 92·4 per cent, and attributes the somewhat lower values of the above authors to their use of the left renal vein which drains the left ovarian or spermatic vein.

When the concentration of amino hippurate or diodrast iodine in the arterial plasma is raised above 5 mg per 100 cc, the amount excreted from each cc of perfusing plasma continues to rise, but not as rapidly as the plasma concentration. In consequence the proportion excreted from the plasma falls progressively.

When amino hippurate exceeds about 20 mg per 100 cc plasma, the amount excreted by the tubules per minute reaches a maximum, and further increases in the plasma concentration affect the excretion rate only by increasing the amount that is excreted by filtration. A similar behavior is shown in the excretion of diodrast and phenol red, and is also shown by the mechanism whereby glucose is reabsorbed *from* the tubular lumina into the circulation As Shannon (1938, 1939) has shown, the mode of transfer of these substances across the tubular wall is consistent with the assumption that the substances combine reversibly with some carrier in the cells of the tubule walls, the carrier combining with the excretory substance on one side of the cell and discharging the substance on the other side. When so much substance is offered to the carrier (high concentration of para-amino hippurate in plasma or of glucose in glomerular filtrate) that the carrier is saturated with the substance, further increase in plasma concentration can cause no further acceleration in tubular transfer and excretion. The process in its kinetics is quantitatively like that of enzymes, which combine with substrates and discharge them as products of enzyme action. That the same carrier in the tubular cells serves to transport more than one substance is indicated by the fact that saturating the carrier with one tubular excreted substance can retard the excretion of another.

The maximum rate of tubular excretion, symbolized as the 'T_m' that can be attained by raising the blood concentration of

para-amino hippurate or diodrast, has been developed by Homer Smith and his collaborators (1951) as a measure of the 'functioning mass' of tubular tissues.

SIGNIFICANCE OF CLEARANCES OF DIFFERENT TYPES OF EXCRETORY SUBSTANCES

The term 'clearance' was introduced (Møller, McIntosh and Van Slyke, 1928) to indicate in a single value, easily visualized as a volume of blood, the relation between the blood concentration of urea and the rate of its excretion. A blood urea clearance of 75 cc per minute, the average for a normal man, indicates that excretion in 1 minute clears the urea from 75 cc of blood. In other words, a volume of 75 cc of blood contains the amount of urea excreted per minute. The clearance concept, first applied to urea, has been fruitfully extended by Homer Smith (1951) and his collaborators to express the relation of excretion rates of other substances to their concentrations in blood (blood clearances) and plasma (plasma clearances). Clearance is calculated as R/B, where R is the excretion *rate* in mg of substance per minute, and B is the *concentration* of the substance in mg per cc *in the blood or plasma*. The normal clearances of five substances, and the parts played by glomerular filtration and tubular excretion, are indicated in Fig. 2.

It will be noted that about 40 per cent of the *urea* filtered in the glomeruli is reabsorbed in the tubules. This reabsorption is attributable, not to active cellular reabsorption, but to back diffusion of the highly diffusible urea from the concentrated tubular urine into the blood. When the urine becomes highly concentrated (urine volume under about 2 cc per minute), the back diffusion of urea becomes even greater than the 40 per cent indicated in Fig. 2. The reabsorption is explainable by the laws of physical diff [illegible]
urea the coeff [illegible]
uniquely low. [illegible]
of the urea diffused back, so that all of that filtered could be excreted. The urea clearance can be pathologically decreased either by decrease in glomerular filtration rate or by an increase in back diffusion through damaged tubules. As will be seen later,

depression of urea clearance in clinical renal failure appears to occur from both causes. When, from any cause, urea clearance in man is depressed permanently below 5 per cent of normal, clinical uraemia impends.

Inulin appears to be freely filtered in the glomeruli, and there

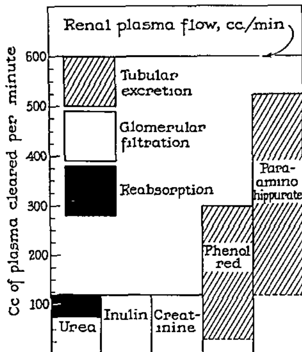


FIG 2. Clearances, by normal men, of substances excreted by glomerular filtration and by tubular excretion

is evidence that as the filtrate passes down *normal* human tubules the inulin in it is neither increased by tubular excretion nor diminished by reabsorption. Hence the plasma clearance of inulin may be assumed under normal conditions to equal the glomerular filtrate. For example, if 600 cc of plasma perfuse the kidney per minute (Fig 2), and 120 cc of plasma water with the inulin dissolved in it are filtered in the glomeruli, the inulin excreted per minute will be the amount contained in 120 cc of arterial plasma, and the plasma inulin clearance will be 120 cc,

equal to the filtrate volume. Hence, inulin clearances are used as measures of the glomerular filtrate, and are often quoted as 'glomerular filtrates'. It seems preferable to report the clearances as such, rather than as filtrate volumes, because in conditions of tubular injury it appears that some back diffusion may occur, so that the assumption that inulin clearance is synonymous with glomerular filtrate may not always hold.

When para-amino hippurate and diodrast are extracted to a normal 90 per cent from the renal plasma, the plasma clearances of these substances approximate 90 per cent of the renal plasma flow, and can serve as a measure of the flow. However, when the tubules are damaged in advanced nephritis, or in the acute renal failure that follows shock or poisoning, as will be shown later, only a small fraction of the PAH or diodrast may be extracted, and their clearances are then much less than the renal blood or plasma flow.

All of the excretory substances represented in Fig. 2 are of the group which the body may be assumed to have no interest in retaining. They are either end-products of metabolism (urea and creatinine) or foreign substances. It is only such substances in general that give clearances that have significance with regard to the functional capacity of the kidney.

For substances like glucose and the amino-acids, which are almost completely reabsorbed to retain their nutritive values, or like the normal electrolytes, that are reabsorbed in the proportions needed to maintain their normal contents in the body, clearance values have no definite significance in terms of functional capacity.

MEASUREMENT OF RENAL BLOOD FLOW

If E mg of an excreted substance are extracted by the kidneys from each cc of renal blood, and the substance is excreted in the urine at a rate of R mg per minute, the volume of blood flowing through the kidney can be calculated as R/E cc per minute. The amount extracted from 1 cc of blood is found as the difference, $A—V$, between A, the concentration of the substance in arterial, and V, the concentration in renal venous blood. The principle is that applied by Fick to measurement of the heart

blood output from the rate of the body's oxygen uptake and the oxygen *A—V* difference. The validity of the calculation of renal blood flow does not depend on the mechanism by which excretion is attained; excretion of any of the five substances represented in Fig. 2 may be used. However, the accuracy of the renal blood flow measurement depends on the percentage error in measurement of the arterio-renal concentration difference, *A—V*, of the substance, and this error is likely to be least when the greatest fraction of the substance is extracted. The principle was first applied by the group at the Rockefeller Institute with urea as the excretory substance (Van Slyke *et al.*, 1934). Only about 10 per cent of urea is extracted, so that a 1 per cent error in blood urea analysis could cause a 10 per cent error in estimated blood flow. Later, creatinine and para-amino hippurate were used in measurements on dogs by the same group (Phillips *et al.*, 1946), and on men, with catheterization of the renal vein, by Warren (1944), Bradley (1948), and others. Except in conditions of tubular injury, extraction of the hippurate is about 90 per cent complete (Fig. 2), so that an analytical error of 1 per cent causes hardly more than 1 per cent error in the measurement of renal blood flow.

As mentioned, the plasma clearance of para-amino hippurate or diodrast can serve as an approximate measure of renal plasma flow when the normal 85–95 per cent of the substance is extracted from the renal plasma. Such use of the diodrast clearance was first proposed by Smith, Goldring and Chasis (Bradley, 1948) for 'the normal human kidney'. Abnormalities may greatly decrease extraction Tubular injury, such as occurs in the post shock condition or from nephrotoxic substances, may decrease the fraction extracted to as little as 2 per cent (Bull, Jockes and Lowe, 1949). Bradley (1948) reports that in both early and late glomerulonephritis the extracted fraction may fall. In such conditions the clearances of diodrast and para-amino hippurate may equal only a small fraction of the actual renal blood flow.

PARALLELISM OF BLOOD FLOW, CLEARANCES, AND OXYGEN CONSUMPTION IN NORMAL KIDNEYS

When the renal blood flow varies, as it may from time to time in the normal kidney, the percentages of urea, creatinine, inulin,

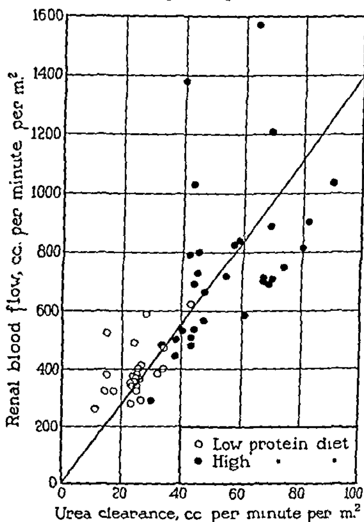


Fig. 3 Direct proportionality of urea clearance to renal blood flow. The high blood flows were induced by very high protein intake. From Van Slyke, Rhoads, Hiller, and Alving (1934)

and para-amino hippurate extracted from the renal blood remain approximately constant. Hence, their rates of excretion vary in direct proportion to the renal blood flow (Fig. 3). Also, the oxygen consumption as shown in Fig. 4 varies in proportion to the renal blood flow (Van Slyke, Rhoads, Hiller and Alving,

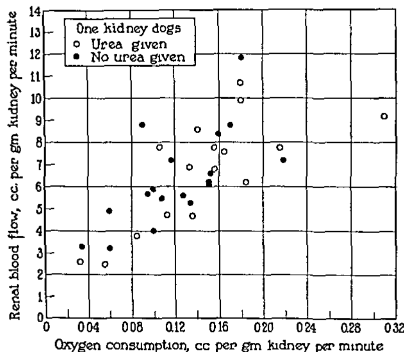


FIG. 4 Direct proportionality of renal oxygen consumption to renal blood flow From Van Slyke, Rhoads, Hiller, and Alving (1934).

1934). In the parallelism between blood flow and oxygen consumption, the kidney appears to be different from the tissues in general. In other parts of the body, if the blood flow becomes slower, tissue oxygen consumption is maintained by extraction of a larger proportion of the oxygen from the blood, and the venous blood becomes darker in color and lower in oxygen content. In contrast to this behavior the blood flow in the normal dog's kidney can be accelerated or retarded by as much as a factor of 3, and still the same proportion of oxygen, averaging

about 10 per cent of the arterial oxygen, is removed from the blood as it passes through the kidney. Even when shock decreases the renal blood flow to as little as 0·3 of normal, the fraction of oxygen removed is not increased (Dole *et al.*, 1946).

The normal parallelism between renal blood flow, oxygen consumption and excretion rates is consistent with the hypothesis that at any moment only a certain proportion of the nephrons is perfused by blood, each nephron with its surrounding tissue extracting a constant proportion of the oxygen from its perfusing blood, and a constant proportion of the plasma's excretory constituents. The phenomenon of perfusion of only part of the glomeruli at a given moment was seen by Richards and Schmidt (1924) under the microscope in frog's kidneys. Khanolkar (1922), from the momentary filtration of haemoglobin and dye in only part of the glomeruli of rabbit's kidneys, *concluded that the same mechanism operated in the rabbit, the circulation opening and closing in the glomeruli, so that at a* given moment only a portion were perfused.

Damage to the nephron cells, such as follows renal ischaemia, destroys these normal parallelisms among the various clearances and the renal blood flow. The clearances become disorganized by tubular inability to reject, reabsorb, or excrete in a normal manner (Phillips and Hamilton, 1948).

MAINTENANCE OF THE BODY'S ACID-BASE BALANCE

One of the functions of the kidney is to maintain a normal pH and a normal bicarbonate reserve in the plasma and in the interstitial fluids. In man and in carnivorous animals there is ordinarily more acid than alkali formed by the metabolism. If the surplus acid in the forms of sulfuric, phosphoric and other acids were not gotten rid of, their accumulation would cause progressive decrease of the bicarbonate reserve of the plasma and interstitial fluids, and a fall of pH to a fatal level. To prevent such a catastrophe the kidney makes use of at least three functions.

1. *Reabsorption of bicarbonate and alkaline phosphate.* The first is reabsorption of bicarbonate and of alkaline phosphate, of the type of Na_2HPO_4, from the glomerular filtrate by the tubules.

The amount of bicarbonate filtered by man per day is estimated at about 400 grammes, and the amount ordinarily in the twenty-four-hour urine is only about 0·2 gramme. Acid phosphate and some of the free carbonic acid are left in the tubules (Sendroy, Seelig and Van Slyke, 1934), and as the result the urine becomes acid. The acidity attained by a solution of either H_2CO_3 or BH_2PO_4 (B represents either Na or K) at the concentration found in the urine, when the bicarbonate and the alkaline phosphate are completely reabsorbed, is indicated by a pH of approximately 5. In fact, a pH slightly below 5 is the lowest reached in the urine. This low pH is attained as a rule only when acid formation in the body is so active that maximal reabsorption of bicarbonate and alkaline phosphate is essential.

2. *Excretion of free acid, or exchange of H^+ for Na^+ and K^+ in the tubules.* A second mechanism for getting rid of acid is either the actual excretion of free acid by the tubular epithelium, or the exchange of hydrogen ions by the epithelial cells for ions of sodium and potassium from the glomerular filtrate in the lumina. It has been demonstrated by Pitts (1946) that when dogs with severe acidosis are infused with large amounts of phosphate the amount of acid phosphate (BH_2PO_4) excreted is greater than the amount filtered in the glomeruli. (The BH_2PO_4 is about one-fifth of the total phosphate filtered at blood pH, the other four-fifths being B_2HPO_4.) Hence, it must be assumed either that the tubules add acid phosphate to the filtrate, or that hydrogen ions are exchanged by the tubular cells for Na^+ and K^+ ions of the B_2HPO_4 in the tubular lumina. This tubular acid excretion is perhaps called into action only when great excess of phosphoric or other acid must be gotten rid of. Under ordinary conditions, and probably under all conditions of clinical acidosis, the acidity of the urine can be accounted for by selective reabsorption of alkaline buffer salts, chiefly $BHCO_3$ and B_2HPO_4.

3 *Excretion of ammonia formed from glutamine and amino-acids in the kidneys.* The third mechanism whereby the kidneys protect the organism from acid intoxication is the formation and excretion of ammonia. In conditions of acidosis the amount of ammonia excreted is usually much greater than the titratable

acid. In diabetic acidosis *ammonia formation is the chief defense* against fatal intoxication. The ammonia is formed in the kidney itself. This was shown by Nash and Benedict (1921) who demonstrated in dogs with acidosis that the ammonia content

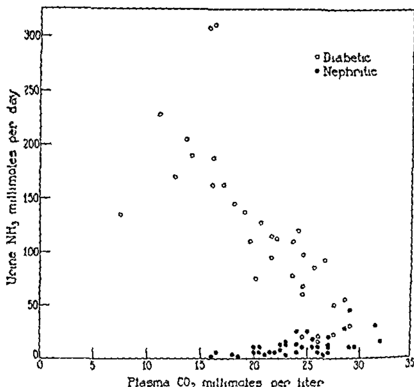


FIG. 5. The diabetic patients show the ability to respond to acidosis, indicated by low plasma CO_2, by great increase in ammonia production. In contrast the nephritic patients show no ability for such response. From records of the Hospital of the Rockefeller Institute.

of the blood in the renal vein was greater than the ammonia content in the arterial blood. If the excreted ammonia had been taken from the blood, the above relation would have been reversed. Fig. 5 shows confirmatory evidence from clinical material that ammonia formation is a kidney function. The stimulus to ammonia formation is acidosis. In the diabetic cases, indicated by the hollow circles in Fig. 5, the excretion of

ammonia rises as the bicarbonate reserve, indicated by the plasma CO_2, decreases. In the nephritic cases, on the contrary, the ammonia excretion actually diminishes as the plasma CO_2 decreases. There is failure of the kidney to respond to acidosis with increased ammonia formation. This failure is attributable to the fact that the diseased kidney loses the ability to form ammonia. It is partly for this reason that acidosis is one of the complications of renal failure.

The source material from which the ammonia excreted by the kidney is formed was at first believed to be urea, then, from the presence of deaminase in kidney tissue, the α-amino nitrogen of amino-acids (Krebs, 1936). The group in our laboratory (Van Slyke *et al.*, 1943) carried out experiments to ascertain in the intact animal whether the ammonia in the urine is formed from urea or from amino-acids or from both. Dogs were subjected to the operation of Rhoads (1934) so that blood could be drawn from the renal vein Thereby it was possible by analyses of arterial and renal venous blood to measure the amounts of urea per minute that were withdrawn from the blood by the kidneys and also the amounts of amino-acids, and to measure the blood flows by the method previously described. For the blood analyses especially accurate techniques had to be devised to measure the changes in urea, amino-acids, and ammonia that occurred as the blood passed through the kidneys. The results showed conclusively that no measurable amount of urea was hydrolysed in the kidneys; all the urea removed from the blood was excreted as urea in the urine. However, measurements of α-amino nitrogen in arterial and renal venous blood plasma of dogs in acidosis showed that the amounts extracted from the blood were quite inadequate to account for the large amounts of ammonia excreted in the urine. It therefore became evident that neither urea nor the α-amino nitrogen of the amino-acids served as the main source of the urinary ammonia.

A search for other possibilities eventually proved that the chief source of urinary ammonia in the dog is the acid amide nitrogen of glutamine. The amide nitrogen of glutamine is easily hydrolysed by the reaction shown in Fig. 6, yielding glutamic acid and ammonia Glutamine had at this time never

been demonstrated in the blood. However, Hamilton (1945) devised a microchemical method by which he was able both to demonstrate and to measure glutamine in blood, and Archibald (1944) developed an independent method based on the use of the enzyme, glutaminase, which confirmed quantitatively the

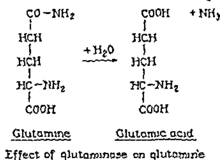


FIG. 6 Formation of ammonia from the acid amide group of glutamine by action of glutaminase.

chemical analyses. By both the chemical and the enzymic method it was shown that glutamine constituted about one-fourth of the total amino-acid content of the plasma, and that the amount of glutamine amide nitrogen removed from the blood plasma by the kidneys was sufficient to account for 60 per cent or more of the ammonia excreted. The ammonia came from the acid amide group of the glutamine, not the α-amino group. Table 2 shows the manner in which change from alkalosis to acidosis increased both the removal of glutamine amide nitrogen from the renal blood and the excretion of ammonia in

TABLE 2 Relation between Rate of Glutamine De-amidation in Renal Blood and Ammonia Excretion in Urine. Dogs

Condition	Urine NH_3—N	Glutamine amide N removed from renal blood
	mg per min	*mg per min.*
Acidosis	0 562	0 33
	0 605	0 39
Alkalosis	0 005	0 02
	0 004	0 02

the urine. In the kidneys of nephritic patients who had lost the greater part of their ability to form ammonia, Archibald (1945) was able to demonstrate that glutaminase was practically absent, whereas normal kidneys contain this enzyme.

In Fig. 7 are diagrammed the apparent mechanisms of

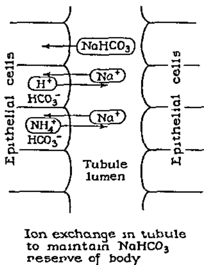


FIG 7 Ion exchange in tubule to maintain bicarbonate reserve of the body.

$NaHCO_3$ reabsorption, and exchange of NH_4^+ and H^+ ions for Na^+ ions of the filtrate to maintain the $NaHCO_3$ reserve of the body. The $NaHCO_3$ (or other bicarbonate salt) absorbed by or formed by ion exchange in the epithelial cells of the tubules is, according to the conception presented, passed back into the blood of the capillary system that surrounds the tubules.

In Fig. 8 the apparent rôles of the different segments of the human nephron are indicated.

UNRELIABILITY OF URINE ANALYSIS FOR DIAGNOSIS OF INTERNAL ACID-BASE ABNORMALITIES

Since the three mechanisms for defense of the body's bicarbonate reserve diagrammed in Fig. 8 ordinarily increase their activity in some proportion to the depletion of that reserve, it

might be expected that unusually low bicarbonate excretion, low urine pH, and high ammonia output would indicate the degree of acidosis in the sense of internal bicarbonate depletion, and that alkalosis would be revealed by the opposite changes in the urine. However, interfering factors are such that observations on the urine can be misleading for diagnosis of the acid-base balance of the body. The reabsorption of $NaHCO_3$ is occasionally so complete in normal conditions that the pH of the urine may be as low as 5, which is the minimum reached in acidosis. The ammonia response to acidosis does not begin at

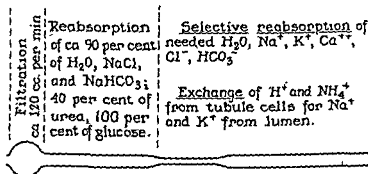


FIG. 8. Diagram of nephron illustrating successive events in the glomerulus, proximal tubule, and distal tubule.

once, and, as shown by Følling (1929), it may require two or three days to reach its maximum. Also, in conditions of renal damage, the kidney may lose most of its ability to form ammonia. Not only in acidosis, but also in alkalosis, diagnosis from the urine may be equivocal.

Such is particularly the case in the type of alkalosis that is perhaps most often encountered in the clinic, viz. that caused by loss of gastric hydrochloric acid by vomiting or lavage. In this condition the loss of gastric HCl may raise the plasma bicarbonate to 50 per cent above normal, a serious alkalosis, and with a high plasma pH. But there is also in this condition a tremendous depletion of the body's NaCl, greatly exceeding the increase in $NaHCO_3$. As a result, the tubular reabsorption of all sodium salts is so stimulated, in the apparent effort to conserve what is left of the body's sodium salt supply, that the urine con-

tains neither chloride nor bicarbonate. The complete reabsorption of the $NaHCO_3$ results in a urinary pH that may be as low as 5 3, despite a plasma bicarbonate over 40 millimoles per liter (Van Slyke and Evans, 1947). A similar condition can be produced by loss of salt through excessive sweating. McCance (1936) has summarized the condition as one in which 'the kidney prefers to regulate the osmotic pressure of the plasma rather than its pH'. Administration of bicarbonate to 'correct an acid urine' in such a case could produce a dangerous increase of the internal alkalosis The required therapy is adequate administration of NaCl and water. When the NaCl and water are replenished, the kidneys quickly excrete the excess bicarbonate (Van Slyke and Evans, 1947).

DIAGNOSTIC APPLICATIONS OF RENAL PHYSIOLOGY

It has been found in glomerular nephritis that as the glomeruli are injured or destroyed the clearances of urea, creatinine, and other excretory substances are diminished. Of these clearances, that of urea is the most easily determined As a measure of *tubular* function, the ability of the kidney to form a urine of high specific gravity is a simple indicator. Addis (1925) has shown that if a normal person is put on a fluid-free diet for twenty-four hours, from after breakfast one morning until the next morning, and the urine for the last twelve hours is collected, the specific gravity will be above 1 026. As nephritis advances the concentrating power of the tubules diminishes, and in advanced nephritis the specific gravity of the urine is but little greater than that of ultra-filtrate of the blood plasma, which is about 1 007. These two measures of renal function, namely urea clearance and the measurement of the [illegible]

dition of the kidneys in the different types of Bright's disease.

Fig 9 shows results obtained in a case of acute glomerular nephritis with recovery (Alving and Van Slyke, 1934). It is seen that the urea clearance returned to normal within two months. The low urinary specific gravity under the conditions of the

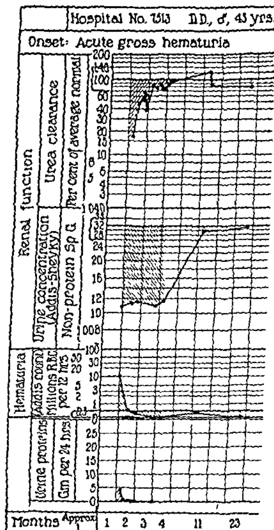


FIG. 9. Course of urea clearance and renal concentrating power of a patient recovering from acute nephritis. For each of these values the horizontal line above the shaded area indicates average normal At the bottom of the chart the first figure after 'months' indicates the number of months the

Addis test, however, showed no increase in these two months, indicating that tubular damage to the concentrating power still remained. Seven months later both functions were entirely normal.

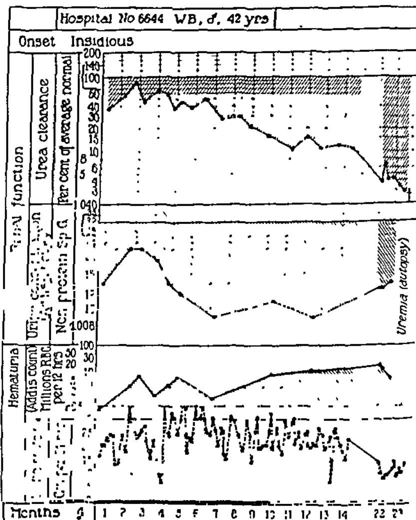
Fig. 10 shows a case (Alving and Van Slyke, 1934) that at first seemed destined to recover like the previous one. However, the initial improvement in function was followed after two months by a progressive fall in the urea clearance and in the concentrating power. After two years the clearance had fallen to about 5 per cent of the normal value and death in uraemia resulted.

gnosis holds regardless of the type of renal damage (Van Slyke and Stillman, 1930). When renal destruction is slow, however, a careful régime may hold off uraemia for months after urea clearance has fallen to 5 per cent. In acute renal failure, as in and after shock, a fall of urea clearance below 1 per cent of normal may be reversible, and followed by recovery (see Fig. 16).

Fig. 11 shows a case (Alving and Van Slyke, 1934) unusual in that over a period of five years the urea clearance was normal following subjective recovery from an acute nephritis, while the specific gravity remained subnormal. The fact that renal pathology continued during this period is shown by the continuance of microscopic haematuria and of proteinuria. Ordinarily in such a patient one would expect a gradual eventual fall in the clearance with termination in uraemia. In this case, however, the outcome was more fortunate. Eventually haematuria and proteinuria disappeared, concentrating power rose to normal, and complete recovery occurred

In 'glomerulonephritis' the tubules as well as the glomeruli are altered, and the change affects tubular function as well as morphology The tubules lose not only their ability to excrete a high specific gravity urine, but they may also lose to varying degrees the power of selective reabsorption, the sodium concentration of the plasma may fall and potassium may rise; the excretion of water may fail to parallel intake, so that variations

in intake that would normally be harmless may cause water intoxication on the one hand or desiccation on the other. And the permeability of the tubules to back diffusion of urea, and



Van Slyke (1934)

perhaps other excretory products, may increase so that unless urine volume is kept high, increased reabsorption of urea adds its effect to decreased filtration in depressing the urea clearance (Van Slyke, 1947). As Marriott has pointed out (1947), such a patient may stave off uraemia by keeping his urine volume up to 2 liters per day, while letting the volume fall to 1 liter may be fatal.

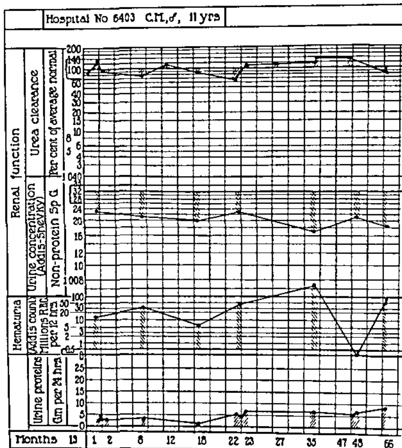


FIG 11 Data from patient with nephritis, showing at first improvement in function, then gradual decline to death in uraemia From Alving and Van Slyke (1934)

GLOMERULAR DESTRUCTION AND DECREASE IN RENAL FUNCTION IN NEPHRITIS

The relation of the percentage of intact glomeruli in the kidneys of nephritic patients to the urea clearance has been studied by Hayman (1934) who used a method by which perfusible glomeruli could be counted in the macerated kidney. The approximate relations obtained by Hayman are indicated in Fig. 12 It

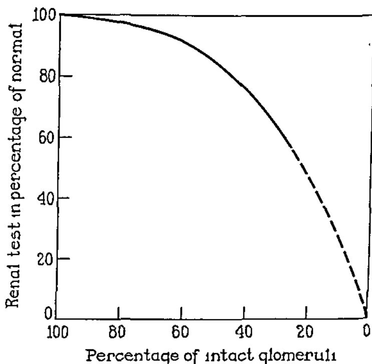


FIG. 12. Relation of glomerular destruction to urea clearance in chronic kidney disease From Hayman and Moritz (1934)

is seen that the loss of the first 50 per cent of glomeruli is accompanied by a fall of only about 10 per cent in the urea clearance. A similar relatively slight effect on urea clearance is obtained when one kidney is removed; the clearance at first falls, but later rises nearly to normal. It appears that destruction or removal of part of the functioning nephrons is followed by hypertrophy, or

increased activity, of the remaining nephrons to such an extent that little loss of function occurs until over 50 per cent of the glomeruli are put out of action. However, as the second 50 per cent of the glomeruli are destroyed a very rapid fall in function follows.

THE SHOCK KIDNEY

Concerning the kidney injured by shock, one could well say that, in the past decade's work that has developed the field, the first word, and the last word, have come from investigators working in London. If I venture to discuss the subject, it is because it has become too important to omit in any review of applied renal physiology.

Acute renal failure, without previous renal disease, occurs in various conditions involving a decrease in the volume of circulating blood Such conditions are severe haemorrhage, trauma, burns, and desiccation. They produce in common the symptoms of collapse, cold extremities, etc., currently considered as the signs of 'shock'.

Anuria during these conditions has probably been noted from the earliest days of clinical observation, but the fact that serious renal injury may remain after recovery from the state of shock was slow in gaining recognition. Forty years ago Rogers (1916–1917) in his classical studies of Asiatic cholera, where anuria accompanies shock from desiccation, found that a considerable number of cases, after their desiccation was corrected by saline infusion, recovered from the shock and the cholera, but not from the renal failure, and died a week later in uraemia. However, it was the observations of Bywaters (1944) and his collaborators in London on the crush syndrome during the first years of the late war that awakened investigators to study what, for want of a better name, we call the 'shock kidney'. It is a pleasure to have this opportunity to acknowledge that work on this subject in the Rockefeller Institute was directly inspired by the reports of the Bywaters group. As the result of studies which they pioneered, clinicians have become alert to this renal syndrome, and I am told by a colleague in a New York hospital that the number of recognized cases of the condition in the hospital at present is likely to exceed the number of all other renal cases.

The First and Second Stages of the Shock Kidney. The development of the shock kidney may be divided into two stages. The first is that of depressed renal function due to circulatory changes, without organic injury to the kidney. If the condition does not

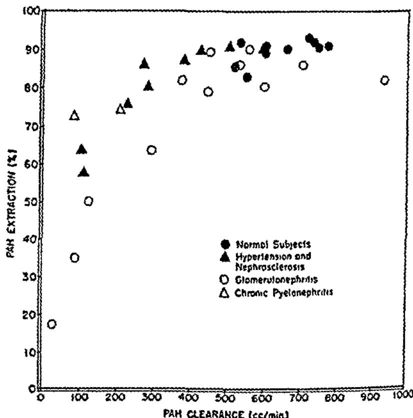


Fig. 13 Decrease in ability of tubules to extract para-amino hippurate from renal blood plasma in advanced nephritis From Cargill (1949).

go beyond this stage, renal function is quickly restored by restoring the general circulation.

The second stage is that of renal damage. If shock is severe and prolonged, restoration of the general circulation may not be accompanied by return of renal function. Anuria, or oliguria with urine of specific gravity 1·008 to 1·010, may persist. Complete or partial renal failure may last until fatal uraemia results,

in a period that may vary from two to twenty days (Lucké, 1946). Or a gradual return of function may occur, with partial or complete recovery. In fatal cases the lesion differs from that found in uraemia from Bright's disease. In the latter the glomeruli are destroyed, with varying changes in the tubules, and the kidneys are often shrunken. In the shock kidney the glomeruli show little change, but gross damage is apparent in the tubules, and the kidneys are usually swollen to much more than

Shock. In Fig. 14 are the immediate effects of progressing haemorrhage on different parts of the renal function (Phillips *et al.*, 1946). Similar changes were observed during shock from muscle trauma. The renal blood flows were calculated from para-amino hippurate extraction and excretion, as previously outlined. The 'filtered fraction', viz. the fraction of plasma water filtered in the glomeruli, was calculated as plasma creatinine clearance/renal plasma flow: this calculation appears to be justified as long as tubular damage does not reach the point where it permits back diffusion of creatinine. The curves in Fig. 14 illustrate the following effects, which were also noted in traumatic shock. (1) General arterial blood pressure may be maintained until a large volume of blood has been lost. In this experiment there was no fall in pressure of the femoral artery until the volume of blood drawn, 44 cc per kilo, was about half that originally present in the dog The maintenance of pressure during such a loss is attributable to peripheral constriction This constriction, while causing some of the evident signs of shock, such as bloodless skin and cold limbs, nevertheless reduces the volume of perfused vascular bed in proportion to the blood loss so that circulation and arterial pressure can be maintained in the heart and brain. (2) The kidneys may not at first take part in the peripheral constriction, but eventually do so. In this experiment the renal blood (plasma) flow remained normal until the blood loss was 25 cc per kilo. With further blood loss a rapid fall in renal blood flow set in. It appeared due to renal constriction, because the femoral blood pressure for the next two hours remained normal, while renal blood flow fell by half. It is as

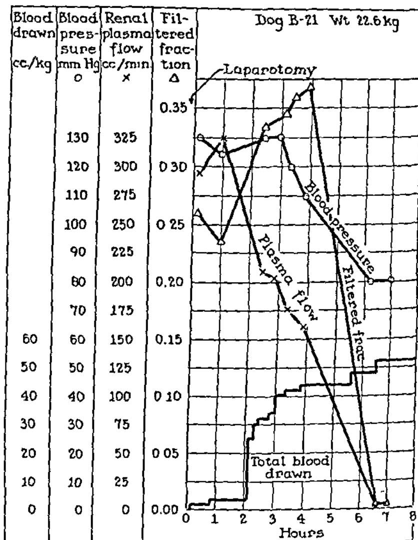
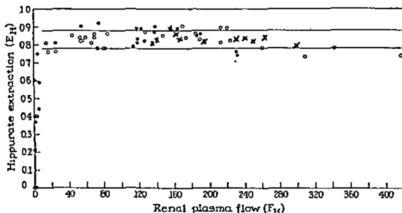


FIG. 14 Renal physiology in shock caused by progressive haemorrhage. From Phillips, Dole, Hamilton, Emerson, Archibald, and Van Slyke (1946).

though the kidney decreased its own circulation in order to maintain that of the heart and brain. Since the normal renal blood flow is about 20 per cent of the cardiac output during rest, there is an obvious advantage to the organism in a mech-

anism for commandeering blood from the kidneys in time of vital need. (3) A compensatory effect in maintaining excretion during hours one to four is seen in the marked increase in the filtered fraction during this period, presumably due to efferent constriction. (4) Eventually the sacrifice of the renal circulation may become complete. In this experiment it occurred after the volume of blood drawn was about 48 cc per kilo. There was a rapid fall in renal blood flow and excretion of creatinine and



hippurate to almost zero. This débacle occurred when the femoral pressure was still over 80 mm, and hence appears to be attributable to renal constriction.

Fig. 15, from experiments like that of Fig 14 with acute shock of gradually progressing severity, shows that extraction of para-amino hippurate from the renal plasma continued to be the normal 80 to 90 per cent until the shock advanced to such a degree that the renal blood flow was less than 10 per cent of normal. Up to this point, it is evident that all the blood was continuing to flow through functioning nephrons. none could have been diverted through non-functioning channels as envisaged in the shunt described by Trueta *et al.* (1946).

In the most advanced stage of the shock there was a marked

fall in the percentage of hippurate extracted. This could be attributed either to restriction of the renal blood to non-excretory channels, or to damage to the tubules, which normally excrete the greater part of the hippurate. Because of evidence, both functional and histological, of tubular damage, we are inclined to the belief that the final decrease in extraction was due to tubular failure, and that the occurrence of the decrease was a sign that the renal condition was passing from the first stage of shock kidney to the second stage, with organic damage to the tubules.

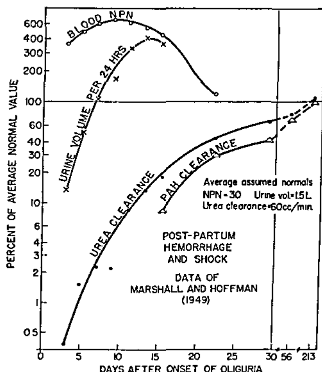


FIG. 16. Chart of patient recovering from haemorrhagic shock. Urea

Functional Picture Persisting after Shock. In Fig. 16 are charted typical data from a case of renal failure following haemorrhagic shock observed by Marshall and Hoffman (1949), who followed the condition from complete anuria to return of normal function attained after several months. The functional data are plotted on a logarithmic scale in order to make visible the early gains in urea clearance, which started at less than 1 per cent of normal. The case illustrates the fact that rise of daily urine volume to the average normal of 1·5 liters, although a favorable prognostic sign, may not indicate that renal excretion of urea, creatinine, etc., has reached similar recovery. In this case on the day the urine volume reached 1 5 liters the urea clearance was only 2 per cent of normal; the urine was hardly more than a plasma ultrafiltrate, with urea concentration about that of the blood instead of the fifty-fold greater concentration that would normally accompany a urine volume of 1 5 liters. Further increase of urine volume to a polyuria, which frequently occurs during recovery, is shown in this case where the daily output in two weeks reached 6 liters. Even at this time the urinary urea concentration was so low that the urea clearance was only 20 per cent of normal. The inability of the kidneys to concentrate the urine indicates lack of tubular concentrating function. Another indication of tubular damage is decrease in the clearance of para-amino hippurate out of proportion to decrease in clearances of mannitol or inulin, which are excreted by glomerular filtration. Normally 70 or 80 per cent of the hippurate excretion is believed to be by tubular excretion, only 20 or 30 per cent by glomerular filtration, and the hippurate clearance is several times the clearances of inulin or mannitol. In the case of Fig 16 the first PAH clearance was no greater than the mannitol clearance (not shown on figure), and in another case was even less than the mannitol clearance.

RÔLES OF DEFICIENT RENAL BLOOD FLOW AND DEFICIENT EXTRACTION OF EXCRETORY PRODUCTS IN EXCRETORY DEFICIT PERSISTING AFTER SHOCK

Until the recent work of Bull, Joekes and Lowe (1950) in the clinics of London University, it was uncertain whether, in cases

fall in the percentage of hippurate extracted. This could be attributed either to restriction of the renal blood to non-excretory channels, or to damage to the tubules, which normally excrete the greater part of the hippurate. Because of evidence, both functional and histological, of tubular damage, we are inclined to the belief that the final decrease in extraction was due to tubular failure, and that the occurrence of the decrease was a sign that the renal condition was passing from the first stage of shock kidney to the second stage, with organic damage to the tubules.

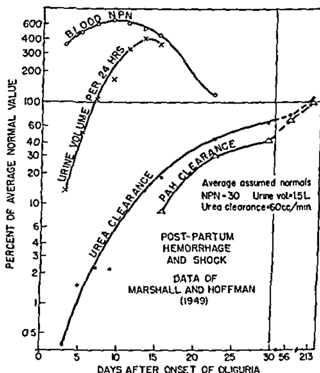


FIG. 16. Chart of patient recovering from haemorrhagic shock. Urea clearance showed progress of improvement from the first day of observation, three days after the onset. However, excretion was not enough to start the

the kidneys found both in autopsies of human cases (Lucké, 1946) and in kidneys of dogs after non-fatal clamping of the renal artery (Roof *et al.*, 1951).

For failure to extract a normal fraction of the urea, creatinine, etc. from the blood that does flow through the kidney, a probable cause is back diffusion through the damaged tubular walls into the blood. That such diffusion can occur through injured tubules was shown in 1929 by A. N. Richards (1929). He watched the nephrons of frogs that had been poisoned with sublethal doses of mercuric chloride. Circulation and filtration in the glomeruli appeared undiminished, but the entire filtrate was reabsorbed, so that none reached the bladder Richards concluded: 'The only explanation which I can reach is that, under these conditions, the osmotic pressure of the blood proteins is unobstructed by the normal qualities of the tubular epithelium and is able to draw all, or nearly all, of the glomerular filtrate back into the bloodstream.' Observations like those of Richards have not yet been made on mammalian kidneys in failure persisting after shock, but the absence of evidence of serious glomerular damage and the presence of gross tubular abnormalities, both structural and functional, lend weight to the probability that a great part of the decrease of extraction is due, not to failure of filtration in the glomeruli of such nephrons as are perfused, but to indiscriminate reabsorption of filtrate, both water and solutes, similar to that observed by Richards through the injured tubules.

This conception is not new. Although Bywaters and his collaborators were not aware of Richards' observation, published in a short address, their recognition of the evidence of tubular malfunction was clear, and one can repeat without change the conclusion reached by Bywaters and Dible (1942) in their paper of 1942 'In those of our fatal cases where adequate biochemical investigation was made, the urine was a dilute fluid showing poor concentration of urea and little reabsorption of chlorides. These characters are those of a poorly concentrated glomerular filtrate and suggest that there is a failure of the tubules to carry out their function. Such a finding agrees with the histological picture, which is in the main tubular. . . . There may be exces-

of renal failure persisting after shock, the diminished clearances of urea and other excretory substances are due to failure of the kidney to extract them from the blood, or to decrease in the volume of the renal blood flow. Thanks to these investigators, it may now be stated that both factors, retarded renal circulation and deficient extraction, contribute in varying but important degrees to the renal failure. Bull, Joekes and Lowe (1950) have used catheterization of the renal vein to ascertain the proportions of urea, creatinine, and para-amino hippurate extracted, and to calculate by the principle of Fick the renal blood flows. An example of their results is shown in Table 3. It is seen that in

TABLE 3. Relative Effects of Retarded Renal Blood Flow and Depressed Extraction on Clearances of PAH and Urea Patient, ten days after onset of Shock Anuria. Data of Bull, Joekes, and Lowe (1950)

	Renal blood flow cc/min.	Fraction of substance extracted		Blood clearance cc/min	
		PAH	Urea	PAH	Urea
	a	*b*	*c*	$a \times b$	$a \times c$
Normal Values	1000	0.900	0.075	900	75
Observed	36	0.021	0.0096	0.75	0.34
Ratio, $\frac{\text{Observed}}{\text{Normal}}$	$\frac{1}{28}$	$\frac{1}{43}$	$\frac{1}{8}$	$\frac{1}{1250}$	$\frac{1}{220}$

this case, ten days after onset from shock, the renal blood flow was only 1/28 of normal, and that from the blood that did flow through the kidneys only 1/8 of the normal proportion of urea was extracted, so that the urea clearance was reduced to 1/220 of normal. Para-amino hippurate extraction was even more decreased, indicating tubular failure.

The cause of the persisting decrease in renal blood flow is uncertain. It may be either persistence in the constriction of renal vessels that is evident during shock, or pressure from oedema of

renal blood flow had nearly returned to normal, but the ability to extract para-amino hippurate from the blood was only one-fourth of normal when the preceding ischaemia had lasted two hours.

FAILURE OF WATER AND ELECTROLYTE REGULATION BY THE SHOCK-INJURED KIDNEY

If the tubules fail to prevent reabsorption of the nitrogenous end-products, we need not be surprised that they fail also in that discriminatory reabsorption of water and of the individual electrolytes required to maintain constant volume and composition of the body fluids. Failure of tubular regulation was noted by Bywaters and his collaborators (Bywaters, 1944; Bywaters and Dible, 1942). It has recently been studied in detail by the late John Lockwood and his collaborators (Randall *et al.*, 1949, 1950) in post-operative conditions, and by Bull, Joekes and Lowe (1950) in post-shock and nephrotoxic conditions. What urine is excreted tends to approach plasma ultrafiltrate in composition. Neither the water nor the individual electrolytes are retained or excreted according to need. In consequence, either excessive retention or depletion of water, sodium, potassium, chloride, or bicarbonate may occur to a dangerous extent, unless administration balances the automatic leakage.

Bull, Joekes and Lowe (1950), in treating 'acute cortical necrosis', caused by shock or renal intoxication, replaced the missing tubular regulation by day-to-day administration of water and individual electrolytes measured to prevent either excess or deficit in the body. Also, a low-protein régime adequate in calories was used to retard accumulation of tissue waste products (Bull, Joekes and Lowe, 1949). With this treatment the mortality, estimated five years ago at 90 per cent (Lucké, 1946), was cut to a small fraction.

INFLUENCE OF ELECTROLYTE ABNORMALITIES ON RECOVERY FROM SHOCK KIDNEY

It appears that the abnormalities of the body's electrolyte and water picture caused by tubular failure may not only injure non-renal organs, as when the heart action is affected by potassium accumulation, or when pulmonary oedema follows excess

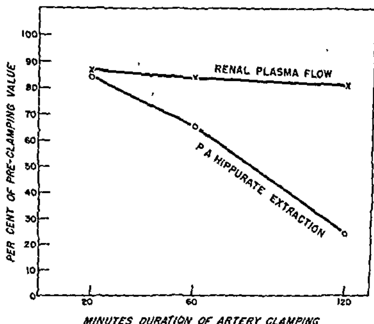


FIG 17. After-effects of clamping the renal artery for periods of from 20 minutes to 2 hours. The measurements of the renal blood flow and extraction of para-amino hippurate were noted approximately 2 hours after the clamps had been removed. The effects of the different periods of renal ischaemia resemble those of mild to moderate renal injury by shock in man. From Phillips and Hamilton (1948).

sive but unselective reabsorption of the glomerular filtrate through the tubules, leakage back into the bloodstream ' What additional facts have since accumulated fit perfectly into this picture.

That deficient tubular function can exist in the presence of normal, or nearly normal, renal blood flow is shown by both clinical and experimental observations. Bull, Joekes and Lowe (1950) found that in patients recovering from shock kidney the renal blood flow was usually restored before the ability of the kidney to concentrate the urine was regained. Phillips and Hamilton (1948) produced a condition resembling moderate damage from shock by clamping the renal arteries of dogs for periods up to two hours. Observations shown in Fig. 17 were taken two hours after the clamps were removed; they show that

structural in basis.' Pigment casts appeared in only two of eleven cases dying within seventy-two hours. From the third day on there appeared the typical necrosis, herniations, and ruptures of the tubules. It appears that the structural changes are secondary to a more subtle disintegration of cellular organization.

RÔLE OF RENAL ISCHAEMIA DURING SHOCK IN PRODUCING ORGANIC INJURY

The question arises: What so injures the kidney during shock that the organ is left in a paralysed or crippled condition? There is evidence that renal ischaemia is the answer. Such evidence, to be convincing, must show that: (1) renal ischaemia occurs during shock, (2) ischaemia by itself can produce the type of renal failure that persists after shock, and (3) ischaemia can produce the type of histological changes observed in post-shock uraemia.

The first investigators to incriminate ischaemia were Scarff and Keele (1943) of the Middlesex Hospital. They did not attack the first point, concerning the occurrence of renal ischaemia during shock, but considered the other two They showed that ischaemia produced in rabbits by clamping the renal artery for one to two hours was followed by a period of renal failure, with nitrogen retention, that could, as in the shock kidney, end in either recovery or uraemia. They also found that tubular lesions developed, similar to those noted in the crush syndrome.

At the same time, the group at the Rockefeller Institute (Van Slyke *et al.*, 1944) were engaged in experiments on dogs that provided the missing link of evidence, viz. proof that renal ischaemia does occur during shock (see Fig. 14), and confirmed the similarity of the functional after-effects of clamping the renal arteries to the after-effects of shock. When recovery occurred, the urea clearance was found to start rising, as shown in Fig. 16, some days before the blood urea began to fall, the behavior resembling that of the human case shown in Fig. 14.

Roof, Lauson, Eder and Bella (1951), continuing the work on the after-effects of ischaemia begun by Phillips and his group, have observed over long periods the after-effects of clamping the renal arteries of dogs for two hours. They found, as after shock

fluid retention, but that the abnormalities may also be part of a vicious cycle retarding or preventing recovery of the kidneys themselves. An example is seen in the shock kidney caused by desiccation in Asiatic cholera. There is a great loss of water and all the electrolytes through the intestine, but relatively more bicarbonate than chloride, because of the alkaline nature of the intestinal secretion. The result is both desiccation and acidosis. Anuria or oliguria with the dilute urine, etc. are typical of the shock kidney results. Sellards and Shaklee (1911) treated seventy-eight cases by infusion of Ringer's solution, containing all the electrolytes except bicarbonate. The desiccation was relieved, but not the severe acidosis, and in twelve cases the renal failure continued and death occurred in uraemia. In seventy-seven other cases adequate sodium bicarbonate to relieve the acidosis was given, and only one death from uraemia occurred. When anuria had been present for a day or less, infusion of bicarbonate was often followed in a few hours by a rapid increase in both volume and urea concentration (e.g. from 0·05 to 1 5 gm per 100 cc) of the urine. Rogers (1916–17), in treating cholera in India, found that by giving adequate amounts of sodium bicarbonate in addition to sodium and calcium chloride solution, the percentage of cases that died later in uraemia was reduced from eleven to three. It is well to emphasize that both Sellards and Rogers followed changes in blood concentration and alkalinity by quantitative methods, and measured their administrations accordingly. It appears that correction of the acidosis was essential to provide a favorable medium for recovery of the damaged tubule cells. It is possible that in part the brilliant therapeutic results of Bull, Joekes and Lowe (1950) were due to restoring a favorable medium for recovery of the tubular cells.

THE SEQUENCE OF POST-SHOCK FUNCTIONAL AND HISTOLOGICAL CHANGES

Failure of tubular function develops before lesions become apparent. Mallory (1947) found no visible changes for eighteen hours after onset of shock anuria in wounded soldiers. 'The initial renal insufficiency is, therefore, functional rather than

Lucké's (1946) cases of 'lower nephron nephrosis' was from carbon monoxide poisoning.

With regard to the third point, the similarity of the lesions developing after renal ischaemia to those developing after shock, Dr. Jean Oliver (Oliver, MacDowell and Tracy, 1951) has added confirmatory evidence to that of Scarff and Keele (1943) and of Badenoch and Darmady (1947) and Darmady (1947). By the courtesy of Dr. Oliver I can use some of his photographs in advance of their publication. Oliver has used for many years a procedure (Oliver, 1939) for isolating and observing separate nephrons from acid-macerated kidneys that reveals with especial clarity the location and nature of structural changes in different segments. He consented to apply this technique to dogs' kidneys that had been removed some days after the animals had been submitted by Phillips and his collaborators (Hamilton *et al.*, 1948; Phillips and Hamilton, 1948) to clamp- [illegible]

The preparations were fixed in formalin and stained with iron-haematoxylin. The nuclei appear as unstained vesicles, due apparently to the extraction of their nucleic acids. The mitochondrial rodlets persist through the preparation, although their form is changed from the long, filamentous structures that are seen in well-fixed histological sections to short, elongated granules, which still maintain their original linear orientation.

In Figs. 19 and 20[1] are shown proximal and distal tubules from a patient that died in uraemia nine days after haemorrhagic shock. In Figs. 21 and 22 are shown similar preparations from the kidney of a dog, of which the renal artery had been clamped for two hours by Phillips.

From examination of these and similar preparations, Dr. Oliver concluded that the lesions developed in dogs' kidneys after one to three hours' clamping were identical with the lesions developed after experimental shock, and that both resembled the lesions found in human cases that developed uraemia after *haemorrhagic or traumatic shock. The most characteristic fea-*

[1] For Figs 19–22 see Plates I–IV following p 396.

in human cases, that several months might be required for complete recovery, and that sometimes renal function, despite clinical recovery, never completely attained its former level. One of the tubular functions that they followed was the maximal tubular capacity to excrete paro-amino hippurate, the 'T_m' of

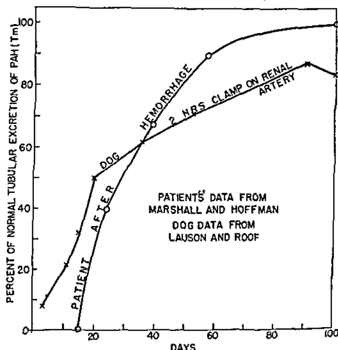


FIG. 18. Comparison of course of rise in tubular excreting power for para-amino hippurate ((T_m) during recovery of patient of Fig. 14, data of Marshall and Hoffman (1949)) with course during recovery of dog from 2-hour clamping of renal artery (data of Roof, Lauson, Eder and Bella (1951))

Homer Smith (1951). In Fig 18 is shown a comparison of the recovery curves of this function in one of the dogs of Roof, Lauson, Eder and Bella and in the patient of Marshall and Hoffman (1949), whose other functional data have been shown in Fig. 16.

Clinical evidence that anoxia can damage the kidney is seen in the occurrence of azotaemia in subjects poisoned with carbon monoxide (Drinker, 1938; Loeper *et al.*, 1941). Also, one of

balances are maintained in the absence of serious cardiac or other non-renal complications, an active life may often be enjoyed until renal destruction has passed the point marked by fall of the urea clearance below 10 per cent of normal, and comfort and some activity may sometimes be maintained until the clearance has fallen below 5 per cent.

REFERENCES

ADDIS, T. A. (1925). A clinical classification of Bright's disease. *J Amer. med Ass* 85, 163

ALVING, A. S, and VAN SLYKE, D. D. (1934). The significance of concentration and dilution tests in Bright's disease *J clin. Invest* 13, 969

ARCHIBALD, R. M (1944). Preparation and assay of glutaminase for glutamine determinations. *J. biol Chem.* 154, 657.

— (1945). Chemical characteristics and physiological rôles of glutamine. *Chem. Rev.* 37, 197.

BADENOCH, W. W., and DARMADY, E M. (1947). The effects of temporary occlusion of the renal artery in rabbits and its relationship to traumatic uremia. *J. Path. Bact.* 59, 79

BORST, J. G. G (1948) Protein katabolism in uraemia. Effects of protein-free diet, infections, and blood-transfusions. *Lancet,* 1, 824

BRADLEY, S E. (1948). *The Pathologic Physiology of Uremia in chronic Bright's Disease* Charles C Thomas, Springfield, Ill

BULL, G. M., JOEKES, A. M, and LOWE, K G (1949). Conservative treatment of anuric uraemia *Lancet,* 2, 229

— JOEKES, A M, and LOWE, K. G. (1950) Renal function studies in acute tubular necrosis *Clin Sci* 9, 379

BYWATERS, E G L (1944) Ischemic muscle necroses Crushing injury, traumatic edema, the crush syndrome, traumatic anuria, compression syndrome a type of injury seen in air-raid casualties following burial beneath debris *J. Amer med. Ass* 124, 1103.

— and DIBLE, J H. (1942) The renal lesion in traumatic anuria *J Path Bact* 54, 111.

CARGILL, WALTER H (1949) The measurement of glomerular and tubular plasma flow in the normal and diseased human kidney. *J clin Invest* 28, 533.

CUSHNY, A R. (1917 and 1926). *The Secretion of Urine.* London

DARMADY, E M. (1947) Renal anoxia and the traumatic uraemia syndrome. *Brit. J Surg* 34, 262

ture of nephron change he considers to be destruction of the basement membrane of the tubule, seen in all four of the preparations, with breaks in the tubule wall which often open directly into veins. This is the lesion first noted by Dunn, Gillespie and Niven (1941) in two cases of the crush syndrome.

In kidneys injured by nephrotoxic poisons, although the functional effects appear the same, Oliver, MacDowell and Tracy (1951) found that the characteristic lesion was different. Instead of part of the nephrons, all were affected, and the characteristic lesion was degeneration of the epithelium rather than the disruptive break, although the latter also might occur.

TUBULAR FAILURE IN CHRONIC RENAL DISEASE

The success of the physiological handling of a primarily tubular condition, exemplified by the work of Bull, Joekes and Lowe (1950), may well provide a model for application of the same principles to such conditions as 'glomerular' nephritis and polycystic kidneys, where tubular failure plays a part that is perhaps not adequately respected. Because of it, excretion of waste products requires extra urine volume to prevent their reabsorption, the relations of sodium, potassium, chloride, bicarbonate, calcium, and phosphate in plasma and interstitial fluids may disintegrate, and inability to form ammonia favors the development of acidosis Acidosis appears to add to the difficulties of an already damaged kidney. I have, for example, seen apparently terminal anuria and coma relieved by a moderate infusion of sodium bicarbonate in a polycystic patient who, under electrolyte and water regulation, afterwards had 15 to 20 per cent of normal urea clearance, and was symptom-free and able to carry on normal activities. In this case the serum potassium could vary with the régime from 3 milliequivalents per liter, where the symptoms of familial paralysis may become evident, to 9 milliequivalents, nearly enough to affect the electrocardiagram, and blood specific gravity went up and down with variations in fluid intake. Treatment of the tubular factor in chronic renal disease has in the past, perhaps, been too often applied only when symptoms became acute, rather than by preventive régime. Provided nutrition (Borst, 1948) and the water and electrolyte

McCance, R. A. (1936). The Goulstonian Lecture. *Lancet*, **1**, 827.
— and Widdowson, E. M. (1937). Alkalosis with disordered kidney functions. *Lancet*, **2**, 247.
Miller, B E, and Winkler, P. W. (1927) Renal excretion of endogenous creatinine in man. Comparison with exogenous creatinine and inulin. *J. clin. Invest.* **17**, 31.
Møller, E., McIntosh, J. F., and Van Slyke, D. D. (1928). Studies of urea excretion. ii. Relationship between urine volume and the rate of urea excretion by normal adults. *J. clin. Invest* **6**, 427.
Nash, T. P., Jr, and Benedict, S. P. (1921). The ammonia content of the blood, and its bearing on the mechanism of acid neutralization in the animal organism. *J. biol. Chem.* **48**, 463.
Oliver, Jean (1939). *Architecture of the Kidney in Chronic Bright's Disease.* London and New York
Oliver, J., MacDowell, M., and Tracy, A. (1951) The pathogenesis of acute renal failure associated with traumatic toxic injury, renal ischemia, nephrotoxic damage, and the ischemuric episode. *J. clin Invest.* **30**, 1305
Phillips, R. A., Dole, V. P., Hamilton, P. B, Emerson, K., Jr., Archibald, R. M, and Van Slyke, D. D (1946). Effects of acute hemorrhagic and traumatic shock on renal function of dogs. *Am. J Physiol.* **145**, 314
— and Hamilton, P. B (1948). Effects of 20, 60, and 120 minutes of renal ischemia on glomerular and tubular function. *Am. J Physiol.* **152**, 523
Pitts, R F., and Lotspeich, W. D. (1946). Factors governing the rate of excretion of titratable acid in the urine *Am J. Physiol.* **147**, 481.
Platt, R. (1950). Sodium and potassium excretion in chronic renal failure. *Clin. Sci* **9**, 369.
R[illegible]

Surgery, **28**, 182.
Rhoads, C P. (1947). A method of explantation of the kidney. *Am J Physiol.* **109**, 324
Richards, A. N (1929) Direct observations of change in function of the renal tubule caused by certain poisons *Trans Ass Am Physicians*, **44**, 64.
— (1938) Processes of urine formation (Croonian Lecture) *Proc roy Soc. B* **126**, 398
— Bott, P A, and Westfall, B B (1938) Experiments concerning the possibility that inulin is secreted by the renal tubules. *Am J. Physiol.* **123**, 281.
— and Schmidt, C F. (1924). Demonstration of intermittency of perfusion through nephrons of frogs *Am J. Physiol* **71**, 178.
Rogers, L (1916–17). Further work on the reduction of the alkalinity of the

DOLE, V. P. (1943). Back diffusion of urea in the mammalian kidney. *Am. J. Physiol.* 139, 504

— EMERSON, K., Jr., PHILLIPS, R. A., HAMILTON, P. B., and VAN SLYKE, D. D. (1949). The renal extraction of oxygen in experimental shock. *Am. J. Physiol.* 145, 337.

DRINKER, C. K. (1938). *Carbon Monoxide Asphyxia.* Oxford University Press, London.

DUNN, J. S., GILLESPIE, M., and NIVEN, J. S. F. (1941). Renal lesions in two cases of crush syndrome. *Lancet*, 2, 549

FOLLINO, A. (1929). On the mechanism of ammonium chloride acidosis. *Acta med. scand.* 71, 221.

GERSH, I., and STIEGLITZ, E. J (1933–4). Histochemical studies on the mammalian kidney. I. The glomerular elimination of ferrocyanide in the rabbit, and some related problems. *Anat. Rec.* 58, 349.

HAMILTON, P. B. (1945). Glutamine: a major constituent of free alpha-amino acids in animal tissues and blood plasma. *J. biol Chem.* 158, 397.

— PHILLIPS, R A., and HILLER, A. (1948). Duration of renal ischemia required to produce uremia. *Am. J. Physiol* 152, 517.

HAYMAN, J. M., Jr., and MORITZ, A R. (1934). The disappearance of

Arch intern Med. 71, 675.

KHANOLKAR, V R (1922) Partial activity of the kidney and the 'all or nothing' principle. *J. Path Bact.* 25, 414.

KREBS, H. A. (1936). Metabolism of amino acids and related substances. *Annu. Rev. Biochem.* 5, 247.

LASHMET, F. H., and NEWBURGH, L H (1932). A comparative study of the excretion of water and solids by normal and abnormal kidneys. *J. clin. Invest.* 11, 1003.

LEAF, A., and CAMARA, A. A. (1949). Renal tubular secretion of potassium in man. *J clin. Invest* 28, 1526.

LOEPER, M., COTTET, J., VARAY, A, and TANASESCO, C. I. (1941). L'azotemie expérimentale dans l'intoxication aigue par le gaz d'éclairage. *C R. Soc. Biol*, Paris. 135, 701

99, 371.

637

n traumatic shock.

Am. J. clin Path 17, 427.

MARRIOTT, H. L (1947) Water and salt depletion. (Croonian Lecture) *Brit med J.* 1, 245, 285 and 328

malian kidney. *Am. J. Physiol.* 99,

between urea excretion, renal blood flow, renal oxygen consumption, and diuresis The mechanism of urea excretion. *Am. J. Physiol* **109**, 336, **110**, 387.

VAN SLYKE, D D, STILLMAN, E., *et al.* (1930). Observations on the courses of different types of Bright's disease, and on the resultant changes in renal anatomy. *Medicine*, 9, 257.

— and EVANS, E. I. (1947). The paradox of aciduria in the presence of alkalosis caused by hypochloremia. *Ann. Surg.* **126**, 545.

WALKER, A (1940). Ammonia formation in the amphibian kidney *Am J. Physiol.* **131**, 186.

— BOTT, P. A., OLIVER, J., and MACDOWELL, M. (1941). Collection and analysis of fluid from single nephrons of mammalian kidney *Am. J. Physiol.* **134**, 580

— and OLIVER, J. (1941). Methods for the collection of fluid from single glomeruli and tubules of the mammalian kidney. *Am J. Physiol* **134**, 562.

WARREN, J. V., BRANNON, E. S., and MERRILL, A. J (1944) A method of obtaining venous renal blood in unanaesthetized persons, with observations on the extraction of oxygen and sodium para-amino hippurate. *Science*, **100**, 108.

blood in Asiatic cholera, and sodium bicarbonate injection in the prevention of uremia. *Ann trop Med. Parasit.* **10**, 129

ROGERS, L (1917). The mortality from post-choleraic uremia. A 70 per cent reduction through intravenous injections of sodium bicarbonate. *Lancet*, **193**, 743.

— (1921). *Bowel Diseases in the Tropics*. London.

ROOF, B. S., LAUSON, H. D., EDER, H. A, and BELLA, T. (1951). Recovery of glomerular and tubular function, including para-amino hippurate extraction, following two hours of renal artery occlusion in the dog *Am J. Physiol.* **166**, 666.

SCARFF, R. W, and KEELE, C. A (1943). The effects of temporary occlusion on the renal circulation of the rabbit. *Brit. J. exp Path* **24**, 147.

SELLARDS, A. W., and SHAKLEE, A. O. (1911). Indications of acid intoxica-

SE

SH

14, 403

— (1939) Renal tubular excretion. *Physiol Rev* **19**, 63.

— and FISHER, S. (1938). The renal tubular reabsorption of glucose in the normal dog. *Am J. Physiol.* **122**, 765.

SMITH, H W. (1951) *The Kidney*. New York.

TRUETA, J., BARCLAY, A E, DANIEL, P. M, FRANKLIN, K J, and PRICHARD, M. L. (1946) Renal pathology in the light of recent neurovascular studies. *Lancet*. **2**, 237.

— (1949) *Studies of the Renal Circulation* Oxford.

VAN SLYKE, D. D. (1942) The kinetics of hydrolytic enzymes *Advanc Enzymol* **2**, 33.

— (1947). The effect of urine volume on urea excretion. *J. clin Invest* **26** 1159.

— PHILLIPS, R A., HAMILTON, P. B, ARCHIBALD, R. M, DOLE, V. P., EMERSON, K, Jr., with tech. assist. of STANLEY, E. G., BECKER, W. H, and OSSERMAN, F. (1944) Effect of shock on the kidney. *Trans. Ass Am. Physicians* **58**, 119.

— PHILLIPS, R A, HAMILTON, P. B., ARCHIBALD, R. M, FUTCHER, P H, and HILLER, A (1943) Glutamine as source material of urinary ammonia. *J biol Chem.* **150**, 481

— HILLER, A., and MILLER, B F. (1935a) The clearance, extraction percentage and estimated filtration of sodium ferrocyanide in the mammalian kidney. Comparison with inulin, creatinine, and urea. *Am J. Physiol.* **113**, 611.

— ... The distribution of ferrocyanide

THE IRRELEVANT ANTIGEN

The first difficulty is the deplorably catholic tastes of the antibody-forming mechanism in man and animals. Whatever survival value the possession of such a mechanism has for a species like man—as a means of dealing either with the large molecules or the living parasites that get into the body—the mechanism has developed little discrimination. The world is full of antigens, bacterial or otherwise, and given the chance, the body responds to all of them. In the would-be immunizer's view, the mechanism is as likely to make a wholly irrelevant response to an invading parasite as it is to produce a protective antibody. Secondly, bacteria are bags of antigens. The flagella and cell-walls, and large molecular cell-substances—proteins, carbohydrates, lipids or combinations of them—enzymes and so forth, may all induce antibodies, though only a few of them may be immunizing antigens.

THE SIMPLE VACCINE

None of these irrelevant antigens and antibodies would matter if in every disease it was both sufficient and harmless to inoculate, let us say, whole cultures of bacteria. But this simple procedure is not often successful. For example, the immunizing antigen may be too toxic, as it is with most of the exotoxins, and require detoxifying without destroying its antigenic powers: or the irrelevant antigens and other bacterial substances may be so toxic that the maximum tolerated dose of whole vaccine does not contain enough of the immunizing antigens. Furthermore, it may be impossible to preserve in the artificial antigen the immunizing property of the natural infection. But even when we have reasonable experimental evidence that a killed vaccine is a useful prophylactic, we are usually far from knowing beyond all reasonable doubt that the vaccine really works. A rigorous test of a vaccine in the human herd with proper controls has seldom been made. In spite of a great volume of work it is only during the last few years, for example, that we have good proof of the efficacy of B C.G vaccine, and that mainly in a few rather unrepresentative communities. We are in much the same condition with whooping cough vaccine; there is plenty of

X

Some Aspects of Antibacterial Immunity

A. A. MILES

THE field of antibacterial immunity is enormous and a full treatment would occupy most of the lectures in this series. I propose to consider not so much what has been discovered, as how we may best sift out those facts that more immediately concern the practice of medicine. I may not succeed in simplifying the issues that are actively debated nowadays. But in all debates by specialists, even informed onlookers sometimes find it hard to distinguish the issues that are immediately important to them from those whose applications to medicine are still very much in the womb of time; and it is here that I hope to be useful.

All our thinking about immunity to bacteria starts with the simple and early concept that the specific immunity to a second attack of an infective disease is due to the induction of specific antibodies; and that the function of these antibodies, broadly speaking, is to opsonize the invader and so make it susceptible to the attack of phagocytes, and to neutralize its toxins. Now there is little in subsequent work to suggest that this concept does not embody the most important of the specific defence mechanisms against bacterial infection, and upon this concept we have built up the practice of prophylactic inoculation by vaccines, dead or alive, and various preparations from bacteria; and of antiserum therapy of infections. Nevertheless, it is often difficult to apply.

active in the tissues, and is rightly employed in diagnosis. But it is not a very reliable indicator of immunizing antibody, and by implication, of immunizing antigen. Because, as I have already mentioned, most of the antibodies formed are probably irrelevant. The flagellar agglutinins that arise in the enteric fevers, whose detection is excellent for diagnosis, are wholly unprotective. Much more rewarding is the search for antigens associated with pathogenicity. The classical example of the pneumococcus illustrates nearly all the essential features of the discovery of an immunizing antigen (Table 1). To summarize it

TABLE 1. The Immunizing Antigen of the Pneumococcus

	Form	
	Smooth capsulated	Rough non-capsulated
Virulence	+++	±
Resistance to phagocytosi	+++	±
Constant of specific polysaccharide	+++	±
Immunizing power	+++	±

Capsular substance = specific polysaccharide
= 'aggressin'

Anticapsular antibody (a) opsonizes pneumococci
(b) is protective in man and the experimental animal

briefly, the virulent pneumococcus is capsulated, but in the course of variation it may lose its capsule, and in doing so loses virulence, resistance to phagocytosis in the test tube, and power to immunize as a bacterial vaccine; the restoration of the capsule restores all these lost properties.

The capsular substance is largely a polysaccharide, which characterizes the particular pneumococcal type; and the antibody responsible for immunity to that type is anticapsular; because a protective antipneumococcal serum loses potency when the capsular antibody is removed by adsorption. Capsular polysaccharide is not very toxic but it enhances infection by pneumococci of the same type in immunized animals, presumably by neutralizing antibody that would otherwise be available for opsonization; and finally, the anticapsular antibody is cura-

suggestive evidence that it works in man, but little firm proof. We can in fact seldom depend on making a good field trial; and when a vaccine succeeds in a field trial we don't usually know, as scientists, how to make another equally good batch of vaccine, though we may know, as master chefs know, how to set about cooking up a potent immunizing brew in the media kitchen. There is no doubt that some of the pertussis vaccines recently tested by the Medical Research Council's team in this country are active (Report, 1951); and that certain laboratories seem to have the trick of making good vaccines. But there is as yet no test or set of laboratory tests that will with certainty pick out the winners. There are rough correlations of efficacy in man with mouse protection tests, but only rough correlations; occasionally a good mouse vaccine is poor in man and a poor mouse vaccine good in man; and to change in the mouse test from challenging by intracerebral inoculation of living *H. pertussis* to challenge by the intranasal route is to get in some cases quite different estimates of the potency of the vaccines.

With B.C.G. vaccine we have a test of efficacy in man, namely, the development of tuberculin allergy. And if we are prepared to accept a positive tuberculin reaction as evidence of antituberculous immunity we can judge whether the vaccine was a good one; but even with this, the production of a good B.C G. vaccine today is a triumph more of ritual and faith than of science.

I stress these difficulties to show that good immunizing agents cannot be achieved merely by making a preparation of the organism and answering the question, 'Does it work?' We have to be usually content with only *suggestive* evidence in man, and back up our faith with the most rigorous laboratory search for the immunizing antigens, and the best ways of using them.

THE SEARCH FOR IMMUNIZING ANTIGENS

In a search of this kind, we often take our first hints from the infected animal or man. For instance, recovery from the infection may be accompanied by a rise in antibody to one or more of the bacterial substances. Now antibody response in the natural infection is good evidence that a certain bacterium is

what we mean by that much abused word 'toxin'. As a working definition I would call it a microbial substance that observably damages living organisms in ecologically significant conditions. The operative parts of that definition are 'observably' and 'ecologically significant'. Taking 'observably' first, it is not enough to infer a toxic action in natural infections, because on isolation in a purified state a bacterial substance is toxic. You will find the *reductio ad absurdum* of this argument in an advertisement for a popular digestive tablet, which portrays a stomach in the act of excreting all its hydrochloric acid in one concentrated drop; enough, as the picture rightly indicates, to burn a hole in the carpet. We have a more serious example in the *Clostridia*. The α-toxin of *Cl. welchii* is a lecithin-splitting enzyme and is rightly credited with endowing that organism with its pathogenicity. The saprophyte *Cl. bifermentans* also produces an antigenic lecithinase of the same kind which, if concentrated enough, is toxic; but in no way does it confer any pathogenicity on *Cl. bifermentans*; and we have no right to call it a toxin (Miles and Miles, 1950).

By 'ecologically significant' conditions I mean conditions resembling those in the natural relation of host and parasite. For example, many bacteria, pathogenic or otherwise, produce haemolysins: but there is little evidence that the haemolysins of pathogenic bacteria ever haemolyse the blood during an infection. It may be objected that I am making no allowance for cumulative effects, such as an anaemia that develops during a chronic infection by a haemolytic bacterium. This is true, but until such a cumulative effect is demonstrated, I maintain that there is no evidence of toxic action in ecologically significant conditions.

I am labouring this point because loose arguments about toxins lead us to believe that we know far more about the action of infective microbes than we actually do Moreover, if you give an antigen a bad name by calling it toxic, you automatically give a good name—antitoxin—to the antibodies against it, and may mislead others into believing that such antibodies play a significant part in prophylaxis or cure.

tive in man. The picture is not complete, because, for example, virulence is not wholly a matter of capsulation: but, had not the sulphanilamides and antibiotics intervened, it is a picture that would, I imagine, still suffice today as a basis for the specific treatment of pneumococcal pneumonia.

We must not expect to find a comparison of virulent and non-virulent forms always so rewarding as in the pneumococcus. Virulence determines the multiplication of the bacteria in the body, and may depend on substances quite different from those which induce immunity; in such a case the immunizing antigen may be present in the non-virulent forms as well. Indeed, we successfully immunize cattle with avirulent *Br. abortus* (McEwen, 1940), and Otten found a vaccine made from a strain of low virulence effective in man (Otten, 1941). In other words, the antigens peculiar to the virulent forms may be irrelevant.

THE TOXINS

There is a stronger case for antigens associated with virulence, when these antigens are toxic. Thus the case for immunizing against the classical exotoxins, diphtheria, tetanus, botulinum, the gas gangrene and the scarlet fever toxins, is clear-cut. They obviously play a predominant part in the disease, because the symptoms of intoxication largely mimic the natural disease. Moreover, they are neutralized in the body by antitoxins, and antitoxins are curative. This is not to say that antitoxic immunity is the whole answer to these diseases. It is certainly not so with scarlet fever, where antitoxin treatment may scarcely affect the course of infection by the causative streptococcus; and there are uncomfortable suspicions, especially in malignant forms of the disease, that more than antitoxic immunity is required in diphtheria. But we are fully justified by the results in our preoccupation with the antitoxic immunity to these diseases, whatever antibacterial immunity we should aim at in addition.

CRITERIA OF TOXICITY IN BACTERIA

When a bacterium has no obvious exotoxin, the selection of toxic antigens is less easy, and before passing on to the lesser bacterial toxins, I want to digress for a moment to consider

are effective in the natural disease. Endotoxins cannot be said to mimic it, but that is hardly to be expected, since they have no specific action in the experimental animal; but if we exclude the peculiar features, say, of gonorrhoea, meningitis, cholera, typhoid fever and *Bact. coli* pyelitis as consequences of the variation in the organs particularly attacked by these diseases, there is a residuum of fever and intoxication which might be attributed to endotoxins. Moreover, in some cases the number of dead organisms required to kill the animal is of the same order of size as the number to which a smaller living infective dose has grown to by the time the animal dies. It is reasonable therefore to assume that in such diseases the animal dies of the endotoxin. We are fortified in this belief by the fact that inoculation with the purified endotoxin, which in many of these bacterial species is the O somatic antigen, confers active immunity and that O-antisera confer passive immunity.

It would however be premature to claim this as a justification for selection of endotoxins as immunizing antigens, because these O-antisera, though they protect against the disease, are not markedly antitoxic. Indeed, H. R. Morgan (1941) has shown that endotoxin-antitoxin precipitates are themselves toxic, suggesting that the O-antibody, though it combines with the endotoxin, does not necessarily neutralize it in so doing. There is another piece of evidence that the endotoxins are not necessarily toxic in ecologically significant conditions. In an experiment designed to demonstrate the inheritance of immunity to the endotoxin of the mouse typhoid bacillus, Hill, Hatswell and Topley (1940) bred from mice that survived a large dose of endotoxin. When the offspring were mature, they were in turn tested with endotoxin, and the survivors used for breeding. Table 2 summarizes the results. In ten generations, there was a survivor rate of 70–80 per cent to that dose of toxin; of the tenth generation of a control set of mice not subjected to selection by endotoxin, only 20 per cent survived. Clearly there is a substantial inherited resistance to endotoxin. Yet when the tenth generations of selected and control mice were infected with living mouse typhoid bacilli, the death rate in the two groups was similar. In other words, resistance to the endotoxin

THE ENDOTOXINS

The toxic materials in bacillary bodies have been studied in the proteus, coliform, salmonella and dysentery, brucella and pasteurella organisms, in the cholera vibrio, and in the gonococcus and meningococcus; and though there are many differences, they conform to a general pattern (Burrows, 1951). They are complexes of protein-like material, polysaccharides and lipids. The toxicity is usually associated with P and N-containing components in either the conjugated protein or the polysaccharide part of the complex. The polysaccharide seems to confer the immunological specificity, and the conjugated protein either confers or improves the antigenicity of the whole endotoxin.

There is also a close resemblance in biological properties. In the experimental animal the toxic dosage, symptoms and post-mortem pictures are similar. Endotoxins may produce early pyrexia and then shock, diarrhoea, prostration. There is an early hyperglycaemia and later a hypoglycaemia. After death the chief changes are congestion, oedema and small haemorrhages in the gut, the abdominal viscera and lungs. Some observers argue that endotoxins act chiefly upon the central nervous system, others that they interfere with carbohydrate metabolism. The majority, struck by the vascular changes in life and at death of the intoxicated animal, believe that endotoxins are primarily vascular poisons (Delaunay and Lebrun, 1947; Delauney, Lebrun and Cotereau, 1947). But since most of the vascular changes can be seen only in animals lethally or near-lethally intoxicated by the endotoxin, it is hard to decide whether we are dealing with a laboratory phenomenon, or something that parallels actual infection by endotoxin producers. But there are two facts that suggest a predilection for the blood vessels—endotoxins in small doses readily cause haemorrhages in certain rat and mouse tumours; and they are prominent among the substances that sensitize the vascular system to the violent changes we see in the Shwartzman-Sanarelli phenomenon.

But these facts do not help us to decide whether the endotoxins

plain saline bleb would. The discovery of bacterial hyaluronidases, and the demonstration that certain viruses produced bigger skin lesions when they were injected along with hyaluronidase, led naturally to the idea that the enzyme might be a valuable aid to a bacterium in spreading through the tissues; and the idea that the power to synthesize hyaluronidase in part determines the invasive properties of a bacterium has been accepted by many bacteriologists. Nevertheless, the evidence is still very equivocal. In the first place, though hyaluronidase-production is sometimes associated with pathogenic strains of a bacterial species, the non-hyaluronidase producers among such strains do not lack virulence (Crowley, 1944; Pike, 1948; Russell and Sherwood, 1949). The association of hyaluronidase and virulence may therefore be accidental. It is true that, experimentally, hyaluronidase can enhance some infections by spreading them through the host's body. But it is pertinent to inquire where the spread is taking the microbes. Are they going into susceptible or into bactericidal tissues? Hyaluronidase enhances vaccinia skin infection, presumably because a virus particle is infective wherever it is pushed by the injection. But it may decrease local lesions due to bacterial infection, because the inoculum is spread more thinly over a larger region of skin, and is more readily suppressed by the natural defences of the body (Duran-Reynals, 1935). We must also consider the probable conditions of hyaluronidase action in natural infection. The phenomenon of experimental hyaluronidase enhancement is evident only when the inoculum is injected under pressure. The enzyme appears to have no intrinsic powers of spread, so that we should expect it to spread natural infection only when the inflammatory exudation had built up a certain hydrostatic pressure in the tissue (Hechter, 1946 and 1947). We might expect this in the late stages of an infection, but not perhaps in the earlier stages. The significant conditions for establishing the pathogenic effects of hyaluronidase must firstly include some knowledge of whether the requisite pressure is likely to be engendered, and secondly, whether it is to the bacterium's advantage to be spread into relatively healthy tissue. One possible proof is, of course, immunological. Is antibody to hyaluronidase

and resistance to infection were not connected. Now the endotoxins are present on the surface of the bacteria, and indeed characterize their antigenic behaviour in agglutination. It is

TABLE 2. Generations of Mice tested with *S. Typhimurium* Endotoxin

Generation	Percentage survival of offspring of	
	Survivors of Test	Controls
1	—	67
2	41	43
3	50	28
4	29	8
5	48	17
6	37	4
7	71	20
8	59	17
9	56	9
10	77	19

possible therefore in many cases where anti-endotoxic antibodies are protective, they act merely as opsonins, and not as antitoxins. If this is generally true, the principle of assuming that endotoxic antigens are likely to be immunizing antigens has given us the right answer, but for the wrong reasons.

THE AUXILIARY PATHOGENIC FACTORS

There remain those bacterial products that have no demonstrable toxicity, but which are thought to have a decisive influence on the course of infection because they can in suitable circumstances act upon mammalian tissue or on substances in that tissue. I have already mentioned the bacterial haemolysins. Another outstanding example will serve to introduce the arguments for their relevance to immunity. Many bacteria produce spreading factors, which in all cases appear to be hyaluronidases. Now the hyaluronidases act on animal tissues that contain highly polymerized hyaluronic acid. In the skin, for example, the hyaluronic acid is part of the intercellular substance binding the tissue cells together When saline containing hyaluronidase is injected intradermally under pressure, the bleb formed by injection spreads outward and subsides far more rapidly than a

IMMUNITY TO *STREP. PYOGENES*

The list of antigens in *Strep. pyogenes* is disconcertingly complex even though it includes only the streptococcal antigens for which some relevance in pathogenesis is claimed.

TABLE 4. Antigens of *Strep. Pyogenes*

P	(nucleoprotein)	Shared by other species
M	(protein)	Type specific
T	(protein)	Inter-type specific
C	(polysaccharide)	Species specific
	Erythrogenic toxin	
	Streptolysin-O	
	Leucocidin	
	Fibrinolysin	
	Hyaluronidase	
	Protease	
	Ribonuclease	
	Desoxyribonuclease	

First come the antigens which determine the serology of the coccus. The nucleoprotein P antigen is shared by other species. This fact does not necessarily exclude it as an effective immunizing antigen but there is no positive evidence that it is. The M and T proteins, which are respectively type specific and shared between types, and the C carbohydrate, which characterizes the *Strep. pyogenes* group as a whole, are all features of pathogenic strains. But since immunity to these varieties of *Strep. pyogenes* is type specific, it is clear that neither T nor C antibodies are concerned in it. The M antigen is associated with virulence and M-antibodies are associated with protection (Lancefield and Dole, 1946; Lancefield and Stewart, 1944). It in fact fulfils all the criteria that indicate the fundamental importance of the capsular substance in the pneumococcus (Table 1). As Rothbard (1948) and Rothbard and Watson (1948) in the U.S. have recently shown, the M antigen is associated with virulence, being plentiful in infecting strains from man, and scanty in carrier strains. It confers resistance to phagocytosis on the coccus, and anti-M sera are opsonic in the test-tube, and protective in experimental infection.

The erythrogenic toxin is a true exotoxin and certainly significant in scarlet fever. It is a species specific toxin, so that the

protective? A test of this kind was made by Evans (1943) in experimental *Cl. welchii* infections in the guinea-pig; and here an antihyaluronidase serum was not beneficial.

The case for hyaluronidase as a factor in pathogenesis should, I think, be regarded as unproven. It is unlikely that we shall ever prove it by finding that it is the *one* feature of a bacterium that confers virulence; and in the absence of such proof, the immunological test—does hyaluronidase immunize against infection, or is antihyaluronidase serum protective?—is crucial. The moral of this analysis is that for proof of pathogenic significance it is not enough to find an enzyme in bacteria that happens to have a substrate in the animal body. The microbial chemists are likely in the next few years to characterize a lot of bacterial enzymes and it would be astounding if these had no effect on some component or other of mammalian tissues. Moreover, they will in all probability be antigenic and as soon as a suitable serological test is devised, antibodies to the enzymes will be demonstrable in the infected animal. I suggest that about these, as about many other bacterial antigens already described, we should keep an open mind, admitting them into the ranks of relevant antigens only after severe scrutiny; and in this scrutiny we must bear in mind the relative value of the various types of evidence of pathogenic significance of an antigen, which as a summary of my argument so far are listed in Table 3 in ascending order of importance and value.

TABLE 3. Types of Evidence for implicating a Bacterial Antigen in Immunity

(1) Specific antibody response in natural disease
(2) Association with virulent form of bacterium.
(3) Observable effect on host or host tissues.
[illegible]

(4) [illegible]

Let us now examine three bacteria, *Strep. pyogenes*, *Staph. pyogenes* and *B. anthracis*, in the light of these criteria.

IMMUNITY TO *STREP. PYOGENES*

The list of antigens in *Strep. pyogenes* is disconcertingly complex even though it includes only the streptococcal antigens for which some relevance in pathogenesis is claimed.

TABLE 4. Antigens of *Strep. Pyogenes*

P (nucleoprotein)	Shared by other species
M (protein)	Type specific
T (protein)	Inter-type specific
C (polysaccharide)	Species specific
	Erythrogenic toxin
	Streptolysin-O
	Leucocidin
	Fibrinolysin
	Hyaluronidase
	Protease
	Ribonuclease
	Desoxyribonuclease

First come the antigens which determine the serology of the coccus. The nucleoprotein P antigen is shared by other species. This fact does not necessarily exclude it as an effective immunizing antigen but there is no positive evidence that it is. The M and T proteins, which are respectively type specific and shared between types, and the C carbohydrate, which characterizes the *Strep. pyogenes* group as a whole, are all features of pathogenic strains. But since immunity to these varieties of *Strep. pyogenes* is type specific, it is clear that neither T nor C antibodies are concerned in it. The M antigen is associated with virulence and M-antibodies are associated with protection (Lancefield and Dole, 1946; Lancefield and Stewart, 1944). It in fact fulfils all the criteria that indicate the fundamental importance of the capsular substance in the pneumococcus (Table 1). As Rothbard (1948) and Rothbard and Watson (1948) in the U.S. have recently shown, the M antigen is associated with virulence, being plentiful in infecting strains from man, and scanty in carrier strains. It confers resistance to phagocytosis on the coccus, and anti-M sera are opsonic in the test-tube, and protective in experimental infection.

The erythrogenic toxin is a true exotoxin and certainly significant in scarlet fever. It is a species specific toxin, so that the

antitoxin is suitable for scarlet fever, no matter what the infecting type of streptococcus.

The streptolysin S, properly speaking, need not be considered here because its antigenicity is doubtful. Human sera will neutralize its haemolytic activity, and the neutralizing titre may change with disease. Neutralization, however, appears to depend not on a true antibody, but a lipo-protein component associated with the albumin fraction of sera. In any event, neutralizing sera are not protective.

The streptolysin-O has a greater claim to our regard, because it is a leucocidin, perhaps the main leucocidin elaborated by the coccus, and in purified state it is toxic. Moreover, anti-O sera are to a small extent protective in experimental infections (Todd, 1935).

The fibrinolysin is not itself lytic. It is an enzyme-like substance which activates a normal component of the serum—named plasminogen—to produce the proteolytic enzyme plasmin. It is not related to strain-virulence in the Group A streptococci, and though anti-fibrinolysins arise in the sera of infected persons, they appear to be irrelevant. It is tempting to see in the power to lyse clots—including the defensive clots believed by some to play a large part in localizing infections—the reasons for the invasiveness of the fibrinolytic streptococci. But there is no good evidence that this is so.

As regards hyaluronidase itself, not all types of *Strep. pyogenes* produce it in large amounts, though since antihyaluronidase antibodies are induced by most types of streptococcus the potentiality is probably fairly widespread. But there is little proof of its significance in immunity. I mention the protease and the two nucleases for the sake of completeness. I can suggest no peculiar pathogenic function for the protease. The streptococcal nucleases, however, are capable of thinning out extremely viscous exudates, like those formed in the pleura, which appear to owe their viscosity to high polymers of ribonucleic acids, presumably derived from pus cells. If the viscous exudate is defensive, a nuclease would be a useful weapon for the coccus, just as a fibrinolysin might be against defensive clots or in destroying the fibrin threads that Smith and Wood (1949) suggest give

mechanical support to phagocytes. We do not know: but perhaps it is not unfair to note that both the fibrinolysin and nuclease, under the names of Streptokinase and Streptodornase, are now employed for the removal of necrotic debris from wounds and closed cavities (Tillett *et al.*, 1950)—a nice modern instance of an old saying, where we beat bacterial swords into clinicians' ploughshares.

We can then simplify the picture of streptococcal immunizing antigens, leaving only the M protein, the erythrogenic toxin, and perhaps streptolysin-O; and if we are stringent in our demands for proof we might exclude the streptolysin-O. The rest, however significant antihyaluronidases, antistreptolysins, antifibrinolysins and antiproteases are in elucidating, for example, the relation of *Strep. pyogenes* to rheumatic fever, we can dismiss for the time being as irrelevant antigens.

IMMUNITY TO *STAPH. PYOGENES*

We can make an equally good simplification with what we know of the *Staph. pyogenes* antigens.

TABLE 5 Antigens of *Staph Pyogenes*

Carbohydrate ?	Species specific
Haemolysins,	alpha (Exotoxin)
	beta
	gamma
	delta
Leucocidin	
Fibrinolysin	
Hyaluronidase	
Coagulase	
(Enterotoxin)	

I do not want to do more with the list in Table 5 than note that we still do not know with certainty the antigens responsible for the undoubted bacterial immunity that follows injection of cocci, nor the conditions in which undoubted antitoxic immun- [illegible]

I am chiefly interested in coagulase as an illustration of my prin-

ciples of selection. The association of coagulase with pathogenicity among the staphylococci is indubitable; but then so is the association with α-toxin production and there is a much better case for α-toxin as an active pathogenic factor; the association of coagulase might then be accidental. But Smith and Hale (Hale and Smith, 1945; Smith, Hale and Smith, 1947) produced evidence firstly that the coagulase could coagulate only by activating (or being activated by) a factor in the blood of the host; and that for animal species without the blood factor, the staphylococcus was not very pathogenic. Secondly they showed that in a coagulable serum, a coagulase-producing staphylococcus increased in resistance to phagocytosis presumably because it was covered by a deposit of fibrin. This is an admirable example of a test made in ecologically significant conditions; that is, the fundamental condition of defence by phagocytes is reproduced. It only remains to prove the protective value of anticoagulase sera in order to establish beyond doubt the pathogenic rôle of the staphylocoagulase in infection.

IMMUNITY TO *B. ANTHRACIS*

My third example, *B. anthracis*, illustrates the failure of all the criteria, excepting the fundamental one, 'Does it work?'

TABLE 6. Antigens of *B. Anthracis*

Somatic polysaccharides PS
Capsular polypeptide PP (neutralizes the anthracidal substance in the tissues)

	Induction of antibody	Immunizing power
Somatic PS	+	0
Capsular PP	+	0
Filtrate factor		
Bacillary substance + 'albumin'	0	+
Lesion factors		
Somatic PS + 'globulin'	0	+
Toxic capsular substance	+	0

Of well-described antigens, we have some polysaccharides in the body of the bacillus, and a capsule which, unlike other

bacterial capsules so far studied, consists of a polypeptide. Both these antigens are active in dead bacilli, and both (Table 6) are irrelevant. This is a particularly instructive and cautionary fact because the capsule is associated with the virulent forms of the anthrax bacillus. We should perhaps expect this result from the older work on anthrax, which established clearly that dead virulent bacilli did not immunize, but living avirulent bacilli did. The recent work in France (Grabar and Staub, 1946; Staub and Grabar, 1947) and the U.S (Cromartie *et al.*, 1947) is based on the old observation that oedema fluid from anthrax lesions could immunize. The French workers' fractionation of oedema fluid gave them a complex of a polysaccharide from the body of the bacillus and a protein, presumably from the tissues. The American workers got similar results, in their hands the antigen was associated with globulin components of the fluid. Gladstone (1946, 1948), on the other hand, sought for an immunizing antigen in cultures of the bacillus When plasma is added to the culture medium, during a particular period of growth the culture contains an immunizing substance which is a complex of bacterial substance and serum albumin. This complex, like the oedema factor, behaved as a protective antigen. The discrepancy in the reputed nature of the protein part of these two immunizing factors may be resolved when its exact nature is known; because the association with albumin or globulin may depend on the method of fractionation employed (Heckly and Goldwasser, 1949). Now this combination of a polysaccharide with a protein to form an antigen is a familiar phenomenon in bacteriology, but there are obstacles to this explanation of the immunizing substances from the anthrax bacillus. The immunity is not transferable in the serum of the animal: antibody to the factor does not appear unless the animal is hyperimmunized, and when it does appear, it reacts in no observable way with any component of the bacterium, of plasma or of tissue extracts.

The nature of the immunity too is peculiar. In the immune animal, as in normal, susceptible animals, the capsulated bacilli proliferate up to 4–8 hours, but after this period, bacilli in the immunized animal begin to lose their capsule, and are later killed. This death does not appear to be brought about by

ciples of selection. The association of coagulase with pathogenicity among the staphylococci is indubitable; but then so is the association with α-toxin production and there is a much better case for α-toxin as an active pathogenic factor; the *association of coagulase might then be accidental. But Smith* and Hale (Hale and Smith, [illegible]
produced evidence firstly t [illegible]
only by activating (or being [illegible]
of the host; and that for animal species without the blood factor, the staphylococcus was not very pathogenic. Secondly they showed that in a coagulable serum, a coagulase-producing staphylococcus increased in resistance to phagocytosis presumably because it was covered by a deposit of fibrin. This is an admirable example of a test made in ecologically significant conditions; that is, the fundamental condition of defence by phagocytes is reproduced. It only remains to prove the protective value of anticoagulase sera in order to establish beyond doubt the pathogenic rôle of the staphylocoagulase in infection.

IMMUNITY TO *B. ANTHRACIS*

My third example, *B. anthracis,* illustrates the failure of all the criteria, excepting the fundamental one, 'Does it work?'

TABLE 6. Antigens of *B. Anthracis*

Somatic polysaccharides PS
Capsular polypeptide PP (neutralizes the anthracidal substance in the tissues)

	Induction of antibody	Immunizing power
Somatic PS	+	0
Capsular PP	+	0
Filtrate factor		
Bacillary substance + 'albumin'	0	+
Lesion factors		
Somatic PS + 'globulin'	0	+
Toxic capsular substance	+	0

Of well-described antigens, we have some polysaccharides in the body of the bacillus, and a capsule which, unlike other

the principle of giving at least two immunizing doses spaced so as to take full advantage of the secondary response described by Glenny (1925) whereby the first dose of an antigen conditions the antibody-forming mechanism in such a way that the second dose given after an interval of about four weeks produces an enormous antibody response. We might also consider whether it is worth improving antibody response in the patient by nutrition But here, I think, we should find that since only gross malnutrition interferes with antibody production, we should not achieve much by overfeeding the reasonably fed; though we should probably do well to ensure, particularly in an infected patient, that there was no lack of the water-soluble vitamins.

PASSIVE IMMUNITY

So far I have discussed active immunity. I should like to finish

remove all the unwanted and inactive serum proteins.

We can also ensure that the antibody gets as quickly as possible to all parts of the infected body by giving it intravenously. But since the infected or intoxicated tissues are usually extravascular, we then have the antibody on one side of the capillary endothelium and its destination on the other. Proteins pass slowly and in relatively small amounts through the normal capillary wall, but more quickly when the permeability of the wall is increased as it is, for example, during inflammation.

Fig. 1 summarizes some observations (Miles and Miles, 1943) on the accumulation of *Cl. welchii* α-antitoxin in artificial lesions made by the α-toxin in the muscle of rabbits given antitoxin at varying times after the toxin In lesions one hour old, for example, the permeability of the capillaries has been increased, and there is a rapid build-up of antitoxin from the circulation. But with a more slowly acting toxin like diphtheria toxin or one which, like tetanus or botulinum toxin, has no effect on the capillary wall, neutralization will depend on the normal slow rate of antibody accumulation in the tissues. Now this may not be quick enough in severe infections or intoxications. Anti-

phagocytosis, though phagocytes appear in the lesion. Cromartie and his colleagues describe in tissues and leucocytes an anthracidal factor which appears to decapsulate and then to kill the bacillus. This anthracidal factor is neutralized not only by the non-toxic capsular polypeptide but also by a substance found by Cromartie and others in lesion fluid, which resembles capsular polypeptide, but which is toxic to animal tissues.

As far as the picture can be safely simplified, we have then a normal defence mechanism of the host—not immunological—the anthracidal factor, which during invasion with virulent [illegible]

of the capsular substance, but in some way or other there is better mobilization of the anthracidal substance, and the infection is overcome locally without spread of the invader.

It would be premature, of course, to assume that the interpretation of this immunity will necessarily lie outside the field of orthodox antigen-antibody effects. We are becoming familiar in allergy with antibodies that are impotent in one way or another, and yet are authentic members of the family of antibodies. Moreover, protective antibodies do appear in animals hyperimmunized to anthrax, so that we could postulate an antibody production that is normally never exuberant enough to spill over into the blood stream in detectable amounts but which is in sufficient concentration in the tissues. But it would be equally unwise to be too conservative. Like tuberculosis, anthrax is a peculiar disease, and in both we may find defence mechanisms, even those which are made more effective by specific immunization, which depend more on a true cellular than on a humoral immunity.

The exploitation of immunizing antigens. When, as in these examples, we have identified the immunizing antigen, there remain the problems of using it to the best advantage, problems which I can do no more than mention. We have various adjuvants that increase antigenic potency—the aluminium salts used with diphtheria toxoid are outstanding in this respect—but it is still doubtful how they act. We can, and should, apply

the antibodies are refined by peptic digestion. The antitoxin has of course to go through more than capillary walls in this placental passage, but there is other evidence that the hold-up occurs at the capillary endothelium, because the natural antibody of the same species (i.e. guinea-pig antitoxin in the guinea-pig) is also best in neutralizing diphtheria toxin, when the toxin

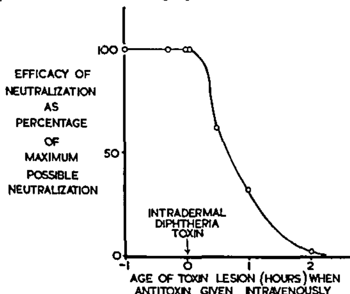


FIG 2. The decline in efficacy of intravenous diphtheria antitoxin administered at varying intervals after intradermal injection of diphtheria toxin in the guinea-pig.

is put into the skin and the antitoxin given intravenously. I do not wish to imply that if these observations prove to be generally applicable all therapeutic horse antisera, and especially refined horse antisera, should be replaced by human antisera. The refined horse antisera are manifestly effective. But there may be diseases and occasions when the greater mobility of the human antibody would prove decisive.

I should perhaps conclude with a justification of what may appear to be a most reactionary view of some recent advances in immunology, in which I have reduced the complexities to a

toxins probably have very little action in *reversing* already established intoxication. They act mainly by stopping further mischief.

In Fig. 2, for example, which summarizes some of our observations on the neutralization of diphtheria toxin in the guinea-pig's skin, by antitoxin in the circulation, it is clear that, in terms of the effective neutralization of the toxin when antitoxin

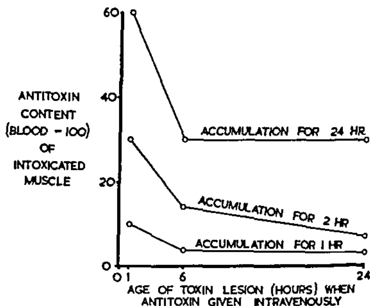


FIG 1. Accumulation of antitoxin in *Cl welchii* α-toxin lesions

is given *before* the toxin, the efficacy of antitoxin given *after* the ...ase, as in others, we must ... of antibody. That must always be a fixed principle of antiserum therapy. But some recent work of Hartley's (1951) suggests that the antibodies in general use are slower than they need be in their passage through the capillary wall. For example, diphtheria antitoxin made in the guinea-pig, and administered to pregnant guinea-pigs, passes much more quickly into the foetus than does diphtheria antitoxin made in the horse: and the rate of passage of both guinea-pig and horse antitoxin is greatly reduced when

the antibodies are refined by peptic digestion. The antitoxin has of course to go through more than capillary walls in this placental passage, but there is other evidence that the hold-up occurs at the capillary endothelium, because the natural antibody of the same species (i.e guinea-pig antitoxin in the guinea-pig) is also best in neutralizing diphtheria toxin, when the toxin

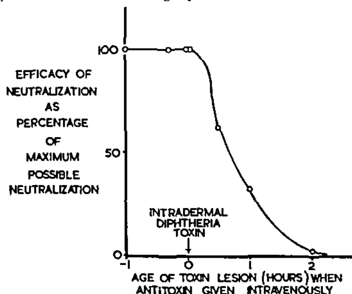


FIG. 2 The decline in efficacy of intravenous diphtheria antitoxin administered at varying intervals after intradermal injection of diphtheria toxin in the guinea-pig.

is put into the skin and the antitoxin given intravenously. I do not wish to imply that if these observations prove to be generally applicable all therapeutic horse antisera, and especially refined horse antisera, should be replaced by human antisera. The refined horse antisera are manifestly effective. But there may be diseases and occasions when the greater mobility of the human antibody would prove decisive.

I should perhaps conclude with a justification of what may appear to be a most reactionary view of some recent advances in immunology, in which I have reduced the complexities to a

toxins probably have very little action in *reversing* already established intoxication. They act mainly by stopping further mischief.

In Fig. 2, for example, which summarizes some of our observations on the neutralization of diphtheria toxin in the guinea-pig's skin, by antitoxin in the circulation, it is clear that, in terms of the effective neutralization of the toxin when antitoxin

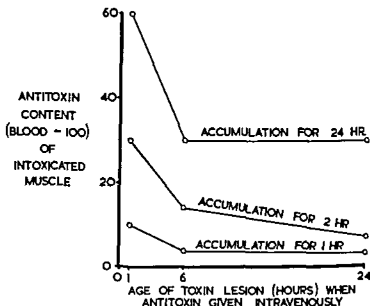


Fig. 1 Accumulation of antitoxin in *Cl. welchii* α-toxin lesions.

is given *before* the toxin, the efficacy of antitoxin given *after* the [illegible] disease, as in others, we must [illegible]ction of antibody. That must always be a fixed principle of antiserum therapy. But some recent work of Hartley's (1951) suggests that the antibodies in general use are slower than they need be in their passage through the capillary wall. For example, diphtheria antitoxin made in the guinea-pig, and administered to pregnant guinea-pigs, passes much more quickly into the foetus than does diphtheria antitoxin made in the horse: and the rate of passage of both guinea-pig and horse antitoxin is greatly reduced when

21.
. infect. Dis. 84, 81.
Path. 28, 57.
A. J., and
HAZLEHURST, G. (1950). *Ann Surg* 131, 12.
TODD, E W (1935). *J. Path. Bact.* 40, 423.

few antitoxic and a few opsonizing antibodies. I do not wish to suggest for one moment that the observations that I have dismissed as irrelevant to antibacterial immunity may not be proved relevant in the future, or that even now they are not of the first importance in bacteriology. They are irrelevant only because the argument has been severely confined to one aspect of immunology. In all conservatism there is an element of laziness, and I have been conservative with the intention of relieving those of my readers who are not professional bacteriologists of some of the burden of bacteriological fact they may feel obliged to carry for their medical thinking. I hope I have been successful.

REFERENCES

BURROWS, W. (1951). *Ann. Rev Microbiol* **5**, 181.
CROMARTIE, W. J. *et al.* (1947). *J. infect. Dis.* **80**, 1, 14, 28, 41, 116, 121.
CROWLEY, N (1944) *J Path. Bact* **56**, 27.
DELAUNAY, A, and LEBRUN, J. (1947) *Ann Inst. Pasteur,* **73**, 555.
— LEBRUN, J, and COTEREAU, H (1947) *Ann Inst Pasteur,* **73**, 565
DURAN-REYNALS, F. (1935). *J exp Med* **61**, 617.
EVANS, D J (1943) *J Path Bact.* **55**, 427
GLADSTONE, G. P. (1946). *Brit J exp Path.* **27**, 394
— (1948) *Ibid* **29**, 379
GLENNY, A T. (1925). *J Hyg, Camb.* **24**, 301.
GRABAR, P, and STAUB, A M (1946). *Ann. Inst Pasteur,* **72**, 534.
HALE, J H., and SMITH, W. (1945). *Brit J. exp. Path.* **26**, 209.
HARTLEY, P (1951) *Proc roy Soc.*, B. **138**, 499.
HECHTER, O (1946). *Science,* **104**, 409.
— (1947) *J. exp Med* **85**, 77
HECKLY, R. J., and GOLDWASSER, E. (1949). *J infect Dis.* **84**, 92.
HILL, A B., HARTSWELL, J. M, and TOPLEY, W. W C. (1940). *J. Hyg., Camb* **40**, 538
LANCEFIELD, R C, and DOLE, V P. (1946). *J exp. Med.* **84**, 449.
— and STEWART, W A. (1944). *J exp Med.* **79**, 79
MCEWEN, A D (1940) *Vet Rec.* **52**, 97.
MEDICAL RESEARCH COUNCIL, Report (1951). *Brit. med. J* **1**, 1464.
MILES, A. A, and MILES, E. M (1943). *Brit J. exp Path* **24**, 95.
[illegible] *Microbiol* **4**, 22
[illegible]
[illegible] *d.* **27**, 111.

PIKE, R. M. (1948). *J infect. Dis* 83, 19.
ROTHBARD, S. (1948). *J. exp. Med* 88, 325.
— and WATSON, R. F. (1948). *J exp Med.* 87, 521.
RUSSELL, B. E., and SHERWOOD, N P. (1949). *J. infect. Dis.* 84, 81.
[illegible]
[illegible] *Path.* 28, 57.
[illegible] A. J., and
HAZLEHURST, G. (1950). *Ann. Surg* 131, 12.
TODD, E. W (1935). *J. Path. Bact.* 40, 423.

XI

Viruses as the Causes of Disease

S. P. BEDSON

VIRUSES are now known to play a part in human and animal pathology which is not surpassed by any other form of microbial agent of disease. The aspects of virus disease that I can most usefully consider concern the ways in which viruses are transmitted from host to host, the routes by which they spread within the host and the manner in which they produce disease. There is something of interest to be said on all these three aspects of virus disease.

MODES OF SPREAD OF VIRUS INFECTION

As a generalization it can be said that infectious diseases are spread by the same means whether they are caused by bacteria, viruses or protozoa. Just as in the case of bacterial diseases there are some virus infections which are spread very largely or even exclusively by contact—lymphogranuloma venereum and rabies are good examples—but the vast majority are spread indirectly through the usual channels—the air, by infection of food, milk and water, by fomites or by insects. The emphasis in the virus group is on air-borne infection and the alimentary tract is relatively unimportant as a portal of entry of virus infection. It is true that examples of virus diseases spread in the latter way are becoming more numerous. There is evidence, for instance, that infective hepatitis may be water-borne (Neefe and Stokes, 1945; Olin, 1947) or even spread at times by food (Read, Bancroft, Doull and Parker, 1946) or by milk (Murphy, Petrie and Work, 1946). And the rediscovery in recent times that those infected with the virus of poliomyelitis excrete the virus in the stools for

weeks or months afterwards, together with its demonstration in

breaks of this disease do occur (Goldstein, Hammon and Viets, 1946) and that it may even, at times, be spread by food, particularly uncooked food (Barber, 1938). It appears, however, that when poliomyelitis is epidemic transmission by air-borne particles is the important mode of spread (McFarlan, 1946; Burnet, 1945) though faecal spread may well be responsible for ensuring survival of the virus between epidemics.

In connexion with the means by which virus diseases are spread the most interesting problems, to my mind, are presented by those diseases in which at first sight there seems to be no evidence that infection is of extrinsic origin. So strong is the suggestion in some instances that the infection arises intrinsically

that this idea of spontaneous and intrinsic origin of viruses has never had many adherents and that a closer examination of apparent examples of its occurrence has provided a more rational and acceptable explanation And although this work is no longer very recent it seems to me worthy of mention because of its extreme interest and because it may have escaped the attention of those who are not primarily concerned with the study of viruses.

The first example which I wish to mention comes from veterinary medicine: swine influenza; we are indebted to the American pathologist Shope for the full understanding of the natural

whilst infection with the bacterium is symptomless; only when acting together do these agents evoke the clinical picture of swine influenza. Although this disease is contagious, the ex-

plosiveness of its outbreak in the late autumn simultaneously at a number of unconnected points and the rapidity with which it spreads through herds and from farm to farm find no conventional explanation. No sporadic cases of swine influenza occur between epidemics nor are carriers of the virus to be found then, and were it not for the brilliant researches of Shope (1943) this yearly outburst of swine influenza might well have been advanced in support of spontaneous generation of the virus. Now we know that swine influenza virus has an important and unexpected host in the swine lungworm. The larvae hatching from the ova laid by adult worms in the lungs of pigs suffering from influenza participate in the virus infection of their host and this infection persists throughout their further development. The first three stages of this larval development are spent in the earthworm and completed when the earthworms are eaten by the pigs and the liberated lungworm larvae reach the lungs by way of the circulation. They bring with them the virus which is not detectable by direct test, but can be unmasked in the pig by the application of a variety of provocative stimuli relatively harmless in themselves. Thus the virus is disseminated widely in swine herds before the onset of swine influenza epidemics but in a form which only becomes unmasked when the infected animals are subjected to the provocative stimulus provided by the wet, cold and inclement weather of late autumn.

The two human conditions to which I wish to make reference in this connexion are herpes febrilis and zoster. As I have said already, the absence of evidence pointing to an external source of infection in those suffering from recurrent attacks of herpes has led to the suggestion that the virus arose spontaneously in the tissues of the sufferer. That the virus is present in the lesions is established and generally accepted, it is only its origin which is in dispute. It is also recognized that herpes can be contagious—the venereal transmission of herpes of the genitals is an example in point—but in the majority of herpetic cases infection by contagion seemed to play no part. The newer knowledge of the natural history of this disease has provided an explanation of this anomaly. The demonstration that adults subject to recurrent herpes possessed herpetic antibody in their blood in con-

siderable amount whereas it was absent from those who never suffered from herpes (Andrewes and Carmichael, 1930; Brain, 1932) suggested that herpetic subjects probably carried the virus. When they acquired the infection which led to the establishment of the carrier state remained unsolved until Dodd, Johnston and Buddingh (1938) showed that vesicular stomatitis in young children was usually due to herpes virus and Burnet and Williams (1939), confirming this, observed that these were primary infections since herpetic antibodies were absent at their inception and only developed as the disease progressed. It seems then that primary infection with herpes virus usually occurs in the early years of life taking the form of an aphthous stomatitis recovery from which leaves a carrier state lasting throughout life. And in those that carry the virus the balance between the resistance of their tissues and the virus may be upset from time to time by one of the many factors known to precipitate an attack of herpes, thus allowing the virus temporarily to gain the upper hand Apparently also if infection with herpes virus is avoided in the first five years of life it is unlikely to be acquired later. Burnet (1945) suggests that the reason for this is that infection is acquired in the home from a parent—the mother most

body, is greatest in those belonging to the lower social strata where the standard of hygiene is lowest. It is not suggested, of course, that primary infection with this virus cannot occur in older children or adults because it does (Scott, Steigman and Convey, 1941) but the chances of acquiring the infection after early childhood are very many fewer and in addition there is the probability that with increasing age the epithelial cells of the mucous membranes become less susceptible to herpes virus.

The epidemiology of zoster presents a very similar problem. The virus, which incidentally has no relationship to herpes virus, can be demonstrated in the vesicle fluid It is of the same size as the virus of varicella, the two viruses cross-react serologically so strongly that with the complement fixation test they are indistinguishable, both produce nuclear inclusions and

neither will infect other animals than man. Infection can be transmitted by intradermal inoculation of vesicle fluid to children who have not previously suffered from either disease and it has been shown experimentally in children that these two viruses cross-protect (Kundratitz, 1925; Bruusgaard, 1932). The evidence suggests, in fact, that the viruses of zoster and varicella are one and the same. When, however, we consider the epidemiology of these two diseases the picture is a very different one. Varicella is a disease primarily of children, it is highly infectious and spreads as varicella. In zoster, as in herpes, infection does not appear to arise extrinsically; in the vast majority of cases zoster is symptomatic, the attack being precipitated by such factors as trauma, X-rays, new growths or poisoning with such things as arsenic; only very rarely does zoster give rise to zoster in those exposed to it, very much more commonly it gives rise to varicella. Zoster is primarily a disease of adults and in a high percentage of cases, some 70 per cent, there is a reliable history of having had varicella earlier in life. These facts suggest that those recovering from varicella continue to carry the virus and that later in life the persisting virus infection can be lit into activity by one of the factors known to precipitate an attack of zoster. In zoster the primary site of virus multiplication, as we know from the work of Head and Campbell (1900), is the ganglia on the posterior nerve roots from which centres the virus spreads centripetally to the posterior, and sometimes the anterior horn of the cord, and centrifugally to the skin where the characteristic vesicular eruption is produced. Whether this means that in the carrier the virus lies dormant in the ganglia or is carried elsewhere, in the blood for example, we do not know, but we have enough information to enable us to explain the natural history of zoster without recourse to such unconventional hypotheses as spontaneous generation. And my purpose in referring in some detail to the natural history of herpes, zoster and swine influenza is to emphasize that closer inspection of these apparent examples of the intrinsic origin of viruses has always provided a more acceptable and biologically sound alternative explanation.

THE SPREAD OF INFECTION WITHIN THE HOST

I should like now to consider briefly the ways in which, after the initial infection, viruses spread within the body of their host. Although in this respect viruses and bacteria often behave alike, there are some differences which are both interesting and important. Whether the virus reaches its host by contact, by being inhaled or by being swallowed, it early meets with susceptible cells which provide the environment necessary for its multiplication. It is in these initial foci of infection that the virus increases in quantity while the disease is incubating and it is from these foci that infection spreads locally or generally. In some cases, such as molluscum contagiosum, the infection never spreads beyond these initial foci and the disease is limited to a few cutaneous lesions with an absence of any general disturbance. Presumably the virus multiplies slowly in those epithelial cells which it has entered and although the virus colony in the cell's cytoplasm comes to assume enormous proportions, the cell seems to suffer little physiological upset and does not die and break down. In the more virulent virus infections, however, death and disintegration of the infected cells occur early, the infection spreads locally and when eventually the local build-up of virus has reached an adequate level it is spread thence to susceptible cells in other parts of the host there to set up those secondary foci of infection which are responsible in major part for the symptoms of the disease. It has long been recognized not only that viruses require cells in which to multiply but that the various species differ in the types of cell that they can utilize for this purpose; they exhibit what we call cell or tissue tropism. The virus of foot and mouth disease is epidermotropic; it multiplies initially in epidermal cells and when it generalizes, as it does through the blood stream, the secondary foci again occur in epithelium. In the past this tissue tropism has, at times, been interpreted too strictly. This has been so, for instance, in the case of the virus of poliomyelitis which was considered to be so strictly neurotropic as to be incapable of multiplying in any but nerve cells, making it difficult to understand how a small inoculum introduced peripherally could provide enough virus

to produce, as it does at times, widespread nerve cell damage or to account for the considerable amount of virus found regularly in the stools. We now know that poliomyelitis virus can make use of the cells of other tissues, including epithelium, for its multiplication in culture of the Maitland type and since also virus has been demonstrated in the tonsil and the wall of the small intestine it seems most probable that in poliomyelitis, as in other virus infections, there is multiplication of the virus at the periphery from which initial foci the central nervous system is later invaded in a small percentage of the infected. On the other hand foot and mouth virus which evinces a predilection for epithelium shows a selectivity even within this range of cells. In the guinea-pig, for instance, this virus will multiply in the epithelium of the soles of the feet, of the dorsum and the edge of the tongue; when injected into the skin of the flanks it fails to produce lesions. And if the virus is injected intramuscularly in a dose adequate to produce a generalized infection it is in the 'pads' of the fore and hind feet and in the epithelium of the dorsum and edge of the tongue that lesions develop. A chance observation made by Mrs. Maitland when working with this virus at the Lister Institute threw some light on this selective location of epithelial lesions. A guinea-pig with a deformity of a hind leg caused by a fracture which resulted in the animal walking on the dorsum of one hind foot, was inoculated in the pads of both hind feet; only in the pad of the normal foot, the one on which it walked, did lesions develop. This suggested to Mrs. Maitland that pressure on the pad epithelium was necessary for the development of foot and mouth lesions. Investigating this possibility she inoculated a normal guinea-pig in the pads of the hind feet and bandaged the feet, protecting them from pressure by a generous layer of cotton wool; lesions failed to develop. Repetition of this basic experiment with variations confirmed that pressure was necessary for foot and mouth virus to develop in the epithelium of the soles of the feet of the guinea-pig. This presumably also explains why lesions of the tongue are confined to its edges and upper surface since it is those parts of the tongue which press against unyielding structures, the teeth and the hard palate. But the full understanding

of this phenomenon escapes us. Possibly the effect of pressure is due simply to slowing of the circulation thus providing the virus in the blood stream with an opportunity of passing through the vessel wall and so of reaching epithelial cells. In any case it is interesting to note that the phenomenon is not peculiar to foot and mouth disease, since it has been shown that it is possible in measles to increase the intensity of the rash in an area of skin simply by strapping the skin firmly before the eruption has occurred. In the two diseases we have just considered, measles and foot and mouth disease, multiplication of the virus in the initial and peripheral lesions is followed by invasion of the blood stream, a viraemia is established giving rise to secondary lesions wherever suitable susceptible cells exist. The same is probably true of the other exanthemata though it would seem that in smallpox the viraemia which is responsible for the skin eruption and the majority of the secondary foci elsewhere is preceded by a transient invasion of the blood stream by virus from the initial foci situated most probably somewhere in the mucosa of the respiratory tract. This virus is rapidly removed from the circulation by reticulo-endothelial cells in the spleen, liver and elsewhere where it multiplies and then reinvades the blood stream in a massive way. This sequence of events in pox diseases has clearly emerged from the studies of mouse-pox (ectromelia) by Fenner (1948) and resembles very closely what happens in typhoid fever. In other virus diseases the spread of infection from the initial lesions takes place by lymphatic channels. This is well exemplified by lymphogranuloma venereum in which condition the virus passes from the insignificant primary lesions on the glans penis or the mucosa of vulva or vagina to the nearest lymphatic glands. The glands involved are commonly the inguinal ones and the bubo so produced is the condition which brings the sufferer to the doctor, the small primary lesion having been overlooked. These modes of spread of infection within the body of the host which we have considered are, however, in no way unusual and are used by bacteria as much as by viruses. A method which is peculiar to viruses is passage along the axis cylinder of nerves The viruses of herpes simplex, zoster and poliomyelitis all spread in this way and though the first

two are, at times, carried by the blood stream, poliomyelitis virus seems to use the axonal mode of spread to the exclusion of all others.[1] Goodpasture was probably the first to provide good evidence of this mode of spread of virus infection. His experiments (Goodpasture, 1925) were made in the rabbit with herpes virus and he relied on the presence of the typical nuclear inclusions produced by this virus as evidence of the presence of virus. He inoculated rabbits in the right masseter muscle with a neurotropic strain of herpes virus, killing the animals as soon as they became febrile and before clinical signs of infection of the central nervous system had appeared. The Vth nerve was then examined histologically and sections of the Gasserian ganglia and of the mid-brain in the neighbourhood of the Vth motor nucleus were also prepared. No lesions were detectable in the Vth nerve whereas the Gasserian ganglion on the inoculated side showed histological change with involvement of the nerve cells. Intracerebral lesions were confined almost exclusively to the nerve cells of the Vth motor nucleus and the number of these cells affected corresponded with what one might have expected had infection come by the axis cylinder of the Vth nerve. And although there was some involvement of neuroglial cells in the neighbourhood of the Vth motor nucleus where the motor fibres of the Vth nerve disperse to enter the nucleus and the myelin sheath thins out, these were few compared with the injury within the nucleus itself. In the case of poliomyelitis there is even better evidence of this axonal mode of spread of the infection. Some years ago Hurst (1930) and Fairbrother and Hurst (1930) showed by histological examination of the central nervous system at different levels between the brain and the nerve cells in the anterior horn of the cord, that in monkeys inoculated intracerebrally with poliomyelitis virus the sequence and disposition of the histological changes indicated that the virus travelled from the brain to the cord by axonal routes. And since that time other investigators have produced confirmatory evidence of a like nature. Recent work has put the matter beyond doubt. Faber, Silverberg and Dong (1950) for instance have

[1] Recent work suggests that a viraemia may occur and play an important part in the early stages of poliomyelitis (Bodian, 1952)

infected monkeys by dipping the cut end of the maxillary division of the Vth nerve into a suspension of poliomyelitis virus. The animals were killed two, three and five days after infection and the Gasserian ganglia and the brain stem, at 0 25 mm. intervals, were examined for histological change; the ganglia were also tested for the presence of virus by subinoculation. Virus reached the Gasserian ganglion by the third day and had spread to the pons and medulla by the fourth or fifth day.

Having reached the pons and medulla the virus spreads upwards to the brain and downwards to the cord, reaching this latter situation by the sixth to seventh day. It is noteworthy that in some of the animals the virus did not spread beyond the Gasserian ganglion. Making use of electron microscopy De Robertis and Schmitt (1949) have demonstrated particles in the neurotubules of sciatic nerves of monkeys infected by nerve-dip. Similar particles could not be demonstrated in preparations of normal sciatic nerves and it was thought that they were probably the virus since they were of the right dimensions and because of the rate at which they travelled along the nerves It has been estimated by Bodian and Howe that poliomyelitis virus [illegible]

observation awaits confirmation the axonal passage of poliomyelitis virus from periphery to centre seems beyond doubt. The importance of this work lies in the fact, as demonstrated by Rous, McMaster and Hudack (1933), that living cells are impermeable to antibody and in consequence poliomyelitis virus in transit from the periphery to the centre by axonal routes is protected against the action of the antibodies. And this no doubt explains the failure of serum therapy to influence the course of the disease however early it may be used in the pre-paralytic stage of the illness.

CHANGES PRODUCED IN THE CELLS AND TISSUES BY VIRUSES

Finally I would like to consider the changes which viruses may produce in the cells which they infect for this, of course, is funda-

mental to an understanding of virus disease. These agents multiply only in intimate association with living cells. There is much evidence that it is in the interior of the infected cell that virus multiplication occurs, though it has recently been suggested by Hoyle (1950) that the final stages in the development of influenza virus may take place at the surface of the infected cell, a [illegible]

indicates that virus multiplication occurs inside cells; the demonstration that so many of the cytoplasmic inclusions produced by viruses are virus colonies and the fact that viruses which have been allowed time to unite with the cells in which they subsequently multiply are protected against neutralizing antibody, support this view. It seems undoubted, also, that the necessity of the living cell for virus multiplication is that these agents are too small to carry the enzymes necessary for an independent existence and that they are therefore forced to make use of the enzyme systems of the cells they parasitize. There seems little doubt that much, if not all, of the damage sustained by the infected cell is due to this borrowed use which the virus makes of the enzymes of its host. The surprising thing, however, is the amount of virus multiplication which the cell will sustain without apparent damage, at least that is what strikes one when following the development of psittacosis virus in tissue culture. My colleagues Barwell, Salaman and Welch have been making a study of this and they have observed cells in which quite appreciable colonies have developed and yet the cells appear in good shape with their mitochondria little, if at all, damaged; and this is with a virus which with further growth and increase in number soon destroys the infected cell. Incidentally their observations (unpublished) do not support Bauer's views on the mode of virus multiplication. He has suggested (Bauer, 1949) that the first result of infection of a cell by a virus is the conversion of the cell cytoplasm into virus cytoplasm which then breaks up into a number of particles which by a process of condensation become elementary bodies. This would imply that there was no increase in the number of virus particles from the

time when the transformed cytoplasm of the cell fragmented, whereas progressive multiplication appears to occur in the case of psittacosis virus. The first readily detectable evidence of multiplication of this virus is the appearance of small aggregates of large virus particles lying in a cell cytoplasm of normal appearance. These virus masses rapidly increase in size until they occupy a large part of the cell and consist of an enormous number of mature virus particles or elementary bodies. This sequence of events suggests multiplication by division, a suggestion which is borne out by the appearance of the virus seen in smears made from the spleens of mice at varying intervals after infection. Another thing that strikes one about the early stages of virus infection as seen in the animal is the absence of any reaction on the part of the tissues to the presence of intracellular virus multiplication. In the lungs of mice which have been infected with psittacosis virus administered intranasally, for instance, cells in the alveolar wall may be seen enlarged and packed with virus and yet the rest of the alveolus has a normal appearance. Only later when these infected cells have broken down and discharged their contents does any response to the infection make its appearance. Then the alveolar walls are thickened and infiltrated with cells some only of which are polymorphs and the alveoli contain a varying admixture of serum, desquamated alveolar cells, red cells and polymorph leucocytes. In virus infections the inflammatory reaction is not primarily to the presence of the virus, it is a response to the cell damage caused by the virus, during the initial virus multiplication the surrounding tissues remain indifferent. This, after all, is what one might expect and is another point in favour of virus multiplication being intracellular.

What is the sequence of changes in a cell infected with a virus? The first effect appears to be one of stimulation for the infected cell commonly increases in size and divides more frequently than normal This leads to hyperplasia of the affected tissue which is well shown in the early stages of the exanthem of smallpox or varicella and the lesions produced in the chorio-allantois of the chick embryo by such viruses as cowpox or herpes. It is true that this hyperplastic effect is usually and rapidly replaced by one of

necrosis when, as the result of the rapid multiplication of virus, the cells die and disintegrate. This occurs in the virus infections I have just mentioned, the pox viruses and herpes, and the hyperplastic papule is replaced by a vesicle which later pustulates. And in some virus infections, of which foot and mouth disease is an example, the march of virus multiplication proceeds at such a tempo that there is not time for hyperplasia to develop and cell destruction dominates the picture almost from the inception of the lesion. In other virus infections, however, death of the cells and necrosis of tissue are either insignificant or entirely absent; the infected cells are stimulated to divide at an enhanced rate, hyperplasia dominates the picture presented by the virus lesion and the result is the formation of a neoplasm. This type of reaction is exemplified by the virus papillomata or human warts, where the resultant growth is non-malignant, and by the filterable fowl tumours such as the Rous fowl sarcomata, many of which are highly malignant. Why certain viruses behave in this way is not known. It is as though the cell by rapid and repeated mitotic division was endeavouring to outstrip virus multiplication, which it possibly does, and so rid itself of its unwanted partner. To suggest that these viruses are lacking in virulence is merely to restate our ignorance. In the majority of virus infections, however, the cells suffer damage which may end in their death and the character of the lesions they produce depends on the balance struck between hyperplasia and necrosis. The cutaneous lesions of variola, varicella, measles and herpes show this admixture; in the lesions produced by this last virus one sees the characteristic ballooning degeneration of the epidermal cells which show varying stages in the development of the inclusions which ultimately occupy the major part of their nuclei. It is the death of these cells which paves the way for vesicle formation.

THE EFFECT ON THE HOST OF THE CELLULAR CHANGES PRODUCED BY VIRUSES

It remains in conclusion to consider briefly the effect on the host of the changes which viruses may produce in cells. Obviously this will depend on the type of change produced and the num-

ber and type of cells affected. In smallpox, measles and varicella in which there is an early dissemination of virus by the blood stream—a viraemia—with the production of lesions in the internal organs as well as in the skin, the general effect on the host is probably the result of the multiplicity of the lesions with breakdown of cells and the consequent liberation of toxic substances, rather than the destruction of any particular cells. There is, of course, the added predisposition to secondary bacterial broncho-pneumonia which occurs particularly in a disease like measles when affecting young children If noble cells are selected for attack by the virus the result of their destruction will depend not only on the number of cells implicated, though this naturally is of prime importance, but also on whether or not the cells destroyed can be replaced. In virus hepatitis, for instance, where in the case of average severity the number of liver cells destroyed is small, complete restoration of the affected liver lobules takes place. Where there is greater destruction of liver cells and particularly where there are repeated attacks or relapses, fibrous replacement of the liver cells may result and a multilobular cirrhosis is produced. Massive destruction of liver cells as in yellow fever would of course be incompatible with life. The death of nerve cells is irreparable and in poliomyelitis, the virus of which disease shows a predilection for the anterior horn cells, although many of the affected cells are capable of recovery, a variable number suffer complete destruction leaving permanent residual paralyses. The virus of lymphogranuloma venereum produces necrosis of the cells in which it multiplies but its choice of cells is the histiocyte and it could be described as mesodermotropic. The virus spreads from the initial lesions in the genital mucosa by the lymphatics and in the secondary stage of this disease the inguinal glands are commonly involved with the formation of a bubo. Though this is a painful and incommoding condition the general effects are minimal, in fact the serious consequences of this disease result from the chronic granulomatous lesions which develop in the genital tract, the rectum or around the anus. These constitute the tertiary stage of lymphogranuloma venereum and may in the long run give rise to rectal stricture, elephantiasis of the

vulva, clitoris or penis or various fistulous states, rectal, recto-vaginal or vesico-vaginal. The serious consequences of lymphogranuloma venereum in a large sense are mechanical and it is only rarely that the virus reaches the blood stream and the infection generalizes.

In the past it has been held that the deleterious effect produced by viruses on their hosts were brought about solely by the indirect means which I have mentioned; any direct toxic action was denied. Recent work necessitates a modification of this view. Possibly as a consequence of the demonstration that rickettsiae were toxic in themselves Rake and Jones (1944) investigated this possibility in the case of the viruses of the psittacosis-lymphogranuloma group and demonstrated clearly that, like the rickettsiae, they were toxic to experimental animals particularly when introduced intravenously. This toxic action is associated with the virus itself, it can be neutralized specifically and is clearly to be distinguished from infectivity. More recently Henle and Henle (1946) have adduced evidence in support of a toxic action by influenza virus and although this evidence is possibly less convincing than in the case of the psittacosis-lymphogranuloma viruses, obviously one has now to bear in mind the possibility of a direct toxic action by the virus in the pathogenesis of virus disease.

REFERENCES

ANDREWES, C. H., and CARMICHAEL, E. A. (1930). *Lancet*, 1, 857.
[illegible]
University Press, Cambridge, Mass.
— and WILLIAMS, S. W. (1939). *Med. J. Aust.* 1, 637.
[illegible] 283.
[illegible] *Pediat.* 12, 96.
[illegible] Vienna.

. Med **91**, 549.
33, 17.
(1946). *J Amer. med. Ass.* **131**, 569.
HOYLE, L. (1950) *J. Hyg* **48**, 277.
J. pub. Hlth. **36**, 169
NEEFE, J. R., and STOKES, J , Jnr. (1945). *J. Amer med. Ass* **128**, 1063.
OLIN, G (1947). *Acta. med. scand.* **128**, Suppl., 196, 381
RAKE, G , and JONES, H. P. (1944). *J. exp. Med.* **79**, 463.
READ, M R., BANCROFT, H., DOULL, J. A , and PARKER, R. F. (1946). *Amer. J. pub Hlth* **36**, 367.
ROUS, P , MCMASTER, P. D., and HUDACK, S. (1933) *Proc. Soc exp. Biol. N.Y.* **31**, 90
SCOTT, T. F. M , STEIGMAN, A J., and CONVEY, J. (1941). *J Amer. med. Ass.* **117**, 999
SHOPE, R. E (1943). *Virus Disease*, Cornell University Press, Ithaca, N.Y.
WYCKOFF, R. W. C. (1951). *Nature, Lond.* **168**, 651.

XII

Antibodies and Immunity to Virus Infection

A. W. DOWNIE

Our knowledge of virus diseases has been greatly extended during the past quarter of a century by the development of new techniques for handling these organisms in the laboratory. The concurrent studies on the nature of immunity to virus infections, however, have to a large extent been directed by what we already know of immunity to bacterial infections. It was at one time held that immunity to virus disease was of a different nature from immunity to bacteria, being essentially based on specific resistance of cells and independent of humoral antibodies. The work of recent years has shown that in acquired immunity to both kinds of infection antibody is of paramount importance. Such differences as have been demonstrated are determined by the exclusively intracellular habitat of the viruses. I may note that what follows concerns virus infections in man and other animals. In plants and bacteria the absence of a vascular system excludes the participation of antibody in resistance to virus attack.

Our views on the significance of antibody in virus infection have been greatly influenced by experimental observations on laboratory animals. We must always keep in mind that in such work the conditions are artificial. In the study of viruses pathogenic for man the viruses used may have been considerably altered by laboratory manipulations, the animals we infect do not suffer from disease due to these viruses under natural conditions, the routes of infection may be unusual and the infecting

virus doses are usually very much greater than those which initiate human disease. None the less from these experiments we have learned something of the conditions in which antibody may effectively prevent infection; and the information and suggestions which these studies provide can and should be tested by observation on virus infections in man.

Immunity to virus infection is a vast subject and here I can only touch upon some aspects of it. After a brief consideration of non-specific factors in resistance I propose to discuss antibody and immunity to virus infection in relation to (*a*) the prevention of infection by convalescent or immune serum, (*b*) correlation of serum antibody levels and resistance, (*c*) estimation of antibody, (*d*) mode of action of antibody, (*e*) development of antibody in relation to clinical course, (*f*) antibody and the duration of immunity and (*g*) subclinical infection.

RESISTANCE TO VIRUS INFECTION INDEPENDENT OF SPECIFIC ANTIBODY

Although the resistance resulting from naturally or artificially acquired virus infection is largely dependent on antibody—the altered serum globulin produced by the tissues in response to virus invasion—there are other less clearly defined factors which help to determine the resistance of persons exposed to an infectious agent for the first time, i e. persons lacking specific antibody. In any community of such persons there is apparently considerable variation in the susceptibility of the individuals within it. The results vary with the nature and virulence of the virus, but in addition to clinical cases, a certain number become infected but show no symptoms—subclinical infections—and in others no invasion of the tissues by the virus occurs. It is in the second and third categories that we presume factors other than [illegible]

readily demonstrated in experimental and natural infections in animals and are apparent in certain human diseases. The child suffers less severely from measles than do adults and yellow fever is a relatively mild disease in children It has been suggested that

this lower susceptibility of the young is related to the greater rate of cellular metabolism but this is hardly a satisfactory explanation. Only suckling mice are susceptible to Coxsackie viruses, while another neurotropic virus produces symptoms only in mature animals.

Mucus and other glandular secretions have been shown to have an inhibitory effect on some viruses, for example influenza virus is inhibited by nasal mucus. Armstrong (1950) has found that infection of mice with Lansing and herpes viruses may be retarded by the addition of mucin to the inoculum. From an analysis of the incidence of poliomyelitis in relation to temperature and humidity, he suggests that these factors are important because of their effect on the mucous secretions of the upper respiratory tract.

The finding of virus inhibitors in bacterial extracts (Ginsberg, Goebel and Horsfall, 1948) suggests that the nature of the bacterial flora of the upper respiratory tract might have some influence in preventing infection in normal individuals. Recent observations have shown that influenza virus and the virus of mumps may be inhibited by the serum of certain individuals in the absence of specific antibody; Wilson Smith and his colleagues have noted variations in this inhibitory property against influenza virus in the same persons at different times (Smith, 1951). My colleagues McCarthy and Germer have been examining normal human sera for inhibitory effect on smallpox virus. Of eleven individuals who had never been vaccinated or exposed to smallpox the serum of five showed appreciable neutralizing activity against the virus. A similar variation in inhibitory activity was shown by sera examined from normal rabbits and guinea-pigs (McCarthy and Germer, 1952). While there is no evidence that such inhibitors in normal secretions, serum and tissue extracts determine resistance to natural infection this is a possibility that at present cannot be excluded from consideration.

The influence of infection of an animal with one virus on subsequent or simultaneous exposure to a second has recently been the subject of a good deal of experimental study. It has been shown that under certain conditions an interference effect on one of the viruses may be demonstrated not only by a different

strain of the same virus but by an unrelated virus. While these studies have provided interesting ideas on the nature of the reaction of cells to virus invasion, the importance of these interference phenomena in human disease is unknown.

The nutritional state of an animal, hormonal influences and the physiological activity of particular cells have also been shown to have some bearing on susceptibility to virus infection under experimental conditions and may well play a part in the field.

THE PREVENTION OF INFECTION BY IMMUNE SERUM

Recovery from infection is usually associated with and probably determined by the development of specific antibody in the infected animal. The immunizing infection may be clinical or subclinical and the immune response can be artificially induced by injections of living, usually attenuated, virus and, in some instances, by concentrated inactive virus vaccines. That antibody is responsible for this actively acquired immunity in man is most readily demonstrable by the protective effect of convalescent serum, or of γ-globulin prepared from it, in measles and certain other diseases. A similar protective or neutralizing effect of convalescent serum can readily be shown in the laboratory in those diseases where the causal virus is pathogenic for laboratory animals. This effect is the basis for some of our diagnostic tests. The protective effect of antibody has also been used, more often in veterinary than in human medicine, by injecting antibody-containing serum at the same time as virulent virus to stimulate active immunity. In such procedures the dose of serum must be carefully chosen so that the virus should not be completely inactivated by the antiserum.

CORRELATION OF SERUM ANTIBODY LEVELS AND RESISTANCE TO INFECTION

If specific immunity against a particular virus is dependent on antibody in the circulating blood or tissue fluid then it should be possible to correlate serum antibody titres with resistance of individuals exposed to the virus. There is not a great deal of evidence of this kind available. Recent observations on influenza

show a general correlation between high antibody levels and resistance to infection. The observations of Maris *et al* (1946) on mumps provide good evidence on this point. They found in a study of groups of children in institutions that of 163 who had complement-fixing antibody in their sera only one contracted mumps, whilst among 283 who had no antibody 56 cases occurred.

There are obvious difficulties in the way of gathering information on the correlation of antibody in man and resistance to specific diseases. Estimations can be made of antibody only against certain viruses. The examination of large numbers of sera may be a formidable task and, even if such examinations were made, subsequent exposure to the virus under consideration might not occur. It is easier to ensure and control exposure to infection in animals, and observations in this field have shown good agreement between the possession of high antibody levels and resistance to infection. Fenner's work on mouse-pox provides a good example of this kind of investigation (Fenner and Fenner, 1949).

THE ESTIMATION OF ANTIBODY

Observations of the kind we are discussing will obviously be of value only if the antibody estimated is active against the virus which is to produce infection and this point requires further examination. There are certain features of our methods for measurement of antibody which may be misleading. Some of the defects of our tests are referable to the nature of the viruses or virus antigens used for test and others are inherent in the kind of test employed. In certain diseases, for example influenza and poliomyelitis, there may be different immunological strains of virus responsible for human outbreaks at different times or places and the strain used for estimation of antibody may not correspond to that responsible for outbreaks under investigation. It seems likely that many observations on antibody titres of human sera to the Lansing strain of poliomyelitis are of little value as an indication of resistance to poliomyelitis strains of different immunological type (see Brown and Francis, 1947). There is the further possibility that the virus strain used in the

laboratory for antibody estimations may, by passage in foreign animal hosts, have lost its original immunological characters.

Techniques used for estimating viral antibody. The tests used fall into two classes: (*a*) *in vitro* tests and (*b*) *in vivo* tests. Of the *in vitro* tests the complement fixation technique is most generally used but the inhibition of haemagglutinins is technically simpler and has been useful particularly in the study of influenza and pox virus antibodies. Both these techniques are convenient but vary in sensitivity; while they serve to indicate an antibody response to virus infection they do not necessarily give an accurate estimate of the power of a serum to neutralize or inactivate the virus. Indeed in the case of vaccinia and variola virus the results of these *in vitro* tests do not closely parallel those obtained by neutralization methods. The *in vivo* tests involve the use of laboratory animals or chick embryos and, while more expensive and time consuming, give a better measure of virus-inhibiting activity of a serum. Of the neutralization techniques, that using the chick embryo chorio-allantois is practicable only against certain viruses, but is a sensitive method of measurement. In variola convalescent sera antibody may be demonstrated by the chick embryo chorio-allantois technique for years after an attack, at a time when *in vitro* tests and neutralization tests in the rabbit give negative results. In general the techniques used for estimating antibody vary in sensitivity and all are probably inadequate for the detection and measurement of antibody. Sera which may be negative for antibody by tests against laboratory strains of virus in unnatural hosts may, in the persons from whom they came, be effective against strains of virus to which these persons may be exposed. In other words, failure to demonstrate antibody by laboratory methods does not necessarily prove its absence.

MODE OF ACTION OF ANTIBODY

Laboratory investigations have largely shaped our ideas on the mode of action of antibody. The experiments of Andrewes (1929) clearly demonstrated that antiserum, to be effective, must make contact with virus before the virus has an opportunity to infect tissue cells. Immune serum injected into the same area of rabbit

skin a few minutes after virus was ineffective in preventing the development of a typical lesion and Rous and his colleagues working with suspensions of cells showed that vaccinia virus within a few minutes of making contact with susceptible cells was thereafter inaccessible to antibody (Rous *et al.*, 1935). This fact has been confirmed for other viruses and by other techniques. Observations of this kind offer an explanation for the failure of immune sera in the treatment of established virus disease. Tissue culture techniques have been used in attempts to analyse the relative importance of cells and antibody in determining the resistance of immune animals. Such experiments have generally shown that virus will grow readily in cells from an immune animal cultured in normal serum provided that the cells have previously been washed free from antibody. On the other hand virus is prevented from growing in cells from a normal animal cultured in serum from an immune animal, provided that the virus makes contact with the serum before it has access to the tissue.

The mechanism of the inhibition of virus by antibody is not known. The antibody is not by itself virucidal, but by combination with virus under suitable conditions antibody renders virus non-infective. It is likely that the mode of action varies with different viruses as immune sera against different bacteria vary in their mode of action. In some virus infections labile factors in fresh serum seem to have a significant effect in reinforcing the action of antibody. Table 1 illustrates this point. It is not known

TABLE 1. The Potentiating Effect of Heat Labile Factors in Human Serum. Neutralization of Variola Virus on the Chorio-allantois by Post-vaccination Serum

Serum dilutions	Per cent neutralization of virus	
	Serum unheated	Serum heated at 58° C.
Undiluted	99	70
1/5	76	0
1/25	26	0

whether antibody combines with virus and prevents it from entering susceptible cells, or whether the virus coated with antibody enters such cells but is prevented from multiplying therein. There have been suggestions, too, from the work of Douglas and Smith (1930), Fairbrother (1933) and others that phagocytic cells may be important in the disposal of virus which has been 'inactivated' by antibody. Antiviral sera are not antitoxic, for viral toxins comparable to bacterial exotoxins have not been demonstrated.

THE DEVELOPMENT OF ANTIBODY AND THE COURSE OF CLINICAL INFECTION

Information on this matter is available only in a few diseases where methods of measuring antibody are practicable In mumps the studies of Henle and his colleagues, to which I will refer again, have shown that antibody may be present in the blood at the time of onset of symptoms. In smallpox antibodies appear towards the end of the first week of illness. Experimental studies of influenza and ectromelia infection in mice have also shown that the appearance of antibody in the blood may pre- [illegible] with the view that [illegible]

In mumps for example the inflammatory swelling of the parotids or submaxillary glands almost certainly represents the reaction to damage and necrosis of cells infected by virus some time before. If, as seems probable, there is blood-stream dissemination of virus during the incubation period, it is likely that the parenchymatous cells of the salivary glands are infected with virus at this time. Within these cells virus will be protected from the action of antibody which may subsequently appear Only when virus proliferation in these cells has proceeded to such a degree as to cause extensive cell necrosis are signs of inflammation likely to arise. Further infection of cells need not occur thereafter in order that signs of disease may develop. Thus antibody present at onset of illness may prevent further infection of cells without inhibiting the inflammatory response to the necro-

sis of cells already infected. In smallpox too, as in ectromelia infection of mice, there is reason to believe that progressive infection of cells stops at the time antibody appears in the circulation, and that the further clinical progress of the disease results from the growth of virus in, and destruction of, cells already infected with virus before the antibody response occurs. On this view death from smallpox may ensue after antibody has appeared in the blood, as a result of widespread infection of cells prior to the development of antibody. This situation has parallels in the field of bacterial infections. Patients may die from tetanus or diphtheria in spite of the injection of large amounts of antitoxin; if lethal amounts of toxin have already been fixed by susceptible tissue the giving of antitoxin is of no avail. In leptospiral infections death may occur in the third or fourth week of illness long after antibodies have appeared in the blood and at a time when leptospira may not be demonstrable in the tissues.

ANTIBODY AND DURATION OF IMMUNITY

Is there any correlation between the persistence of antibody and the duration of immunity after virus infections in man? We may consider briefly two different kinds of infection, namely that exemplified by smallpox, measles or mumps, which are followed by immunity to further attack often lasting for many years, and infections such as epidemic influenza which are followed by a relatively short immunity lasting for months or at most a year or two. The short-lived immunity following influenza is not however always associated with disappearance of antibody from the blood; the difference in the observed duration of immunity between these two types of disease is, in part at least, referable to difference in the nature and location of the infective process.

Smallpox In smallpox there is a generalized infection, the virus being distributed by the blood stream before the commencement of illness to those tissues where further multiplication is to occur. The onset of symptoms coincides with the tremendous growth of virus in the skin and elsewhere and is probably due to the products of cell necrosis. The generalized infection induces a good antibody response on the part of the tissues and in those patients who recover antibody titres are high and antibody may be

detected in the serum for many years. In such individuals exposed to infection a second time, virus would make contact with antibody in the blood before it could reach those tissues where further growth would have to occur before symptoms could be produced.

Influenza. In influenza the immunity is short-lived and second attacks are common. It was at one time thought that second attacks were always the result of infection with a different immunological type of virus. While this undoubtedly happens it has been shown that second attacks due to the same immunological type of virus may occur within a few months (Francis *et al.*, 1944). Moreover antibody resulting from the first infection may still be detectable in the patient's blood at the time of the second attack, thus confirming in man what had been demonstrated in experimental ferret infections by Wilson Smith years ago. In influenza, however, there is no generalization of the infection. The virus implanted on the respiratory mucosa rapidly involves considerable areas of surface epithelium; the symptoms of infection are presumably consequent on necrosis of the infected epithelial cells. As there is no invasion of the blood stream the infecting virus is not exposed to the full concentration of antibody which may be present in the circulation. But antibody to be effective in ensuring protection must make contact with the virus before widespread infection of susceptible cells occurs; and, as Francis has suggested, only specific antibody in the respiratory exudate or secretion could be effective in preventing widespread infection of the mucosa by influenza virus. This view has recently received confirmation by the experimental work of Fazekas de St. Groth and Donnelley (1950) on influenzal infection in mice. Groups of mice were vaccinated by formolized suspensions of viruses A and B given by various routes, eleven days later the mice were challenged by intranasal instillation of active virus Only vaccine prepared from the type used for the subsequent test infection induced immunity and vaccines given by the intranasal route protected better than those injected intraperitoneally. Titration of antibody in the serum and in the bronchial washings showed that in the intranasally vaccinated mice, although the serum antibody might be less, the titre of

antibody in bronchial washings was considerably higher than in mice given the same vaccine intraperitoneally. Fazekas de St. Groth and Donnelley showed in further experiments that if the antibody content of bronchial secretion could be increased by non-specific means a high degree of protection could be obtained. This work seems to offer suggestions for variation in the route of administration of vaccines against influenza.

Herpes Simplex. Infection with the virus of herpes simplex seemed at one time to present a paradoxical state of affairs in that only those persons whose serum possessed antibody suffered recurrent attacks of herpetic infection of the lips or other mucous membranes. The situation has been clarified with the demonstration of the nature of primary herpetic attacks. These usually occur in the first few years of life and may be associated with a good deal of constitutional upset. Indeed the capacity o this virus to produce disease may in the past have been under estimated, for it may be the cause of a fatal encephalitis, o severe and extensive vesicular eruptions in the skin. After th initial infection children become carriers and may in after lif suffer from recurrent attacks of labial herpes, excited sometime by other febrile illnesses. Those adults whose serum possesses n antibody have never been infected with the virus. What then i the nature of the carrier state, and why is the high concentratior of antibody in the blood not effective in preventing recurrence of the latent infection? It should be noted that antibody i effective in preventing severe infections with the virus in the herpetic carrier; widespread infections occur only in occasiona persons who encounter the virus for the first time. The only exception to this that I know is the case of recurrent extensive skin infection recorded in the paper of Boake, Dudgeon and Burnet (1951) Their patient was a diabetic who suffered from atopic eczema and the authors consider that after the first extensive eruption the virus may have persisted in the eczematous areas of skin and was reactivated by the attacks of pneumonia from which the patient suffered. One can only speculate as to the sequence of events which lead to a recurrence of local herpes. The recurrence is perhaps dependent on the tropism of the virus for squamous epithelium and the relatively avascular nature of

that tissue. When the infection is clinically quiescent the virus in infected cells may go on multiplying, but the virus liberated from such cells is for the most part neutralized by antibody in the intercellular fluid. Those conditions which precipitate a clinical recurrence may involve vascular changes leading to alterations in the amount or composition of the intracellular fluid so as to allow virus to escape the neutralizing action of antibody. Once a sufficient number of cells have been infected and destroyed, the local inflammatory reaction which ensues leads to a local concentration of antibody which prevents further extension of infection. Recurrent attacks of herpetic infection are clinically obvious because of their superficial situation. It seems possible that similar local reactivation of other latent viruses might occur in deeper structures in the body but pass undetected. It is known that virus may persist for a long time in immune animals and there have been reports of reactivation of psittacosis in man; latent infections with this virus in apparently healthy birds may become clinically manifest if the birds are badly housed and fed. Zoster may in some cases be a lighting up of a long latent infection with the virus of chickenpox.

SUBCLINICAL INFECTION

The incidence of subclinical infections varies according to the nature and virulence of the infecting virus. For example subclinical infections are common in poliomyelitis and in epidemic influenza, less common in mumps and rare in smallpox Before the era of laboratory research on viruses epidemiological observations indicated that such infections were frequent; and their existence, in influenza for example, can readily be confirmed by detecting the development or increase of antibody in the blood of healthy persons exposed to infection. In individuals who have been infected with a particular virus the subclinical nature of a second or subsequent attack might be determined by the rapid production of antibody by tissues tuned for the task by previous experience. The antibody so produced might prevent the virus from invading a sufficient number of susceptible cells to produce symptoms. But subclinical infection may occur in persons exposed to a virus for the first time. In mumps for example about

a third of all first infections are subclinical and the observations of Henle *et al.* (1948) are of interest in this regard. A group of fifteen children who had never suffered from mumps and whose sera contained no antibody were infected with mumps virus in chick amniotic fluid. Of the infected children seven developed clinical mumps but eight never showed signs of their infection. However, it was found by examining saliva repeatedly from these eight children that virus was excreted by six of them between the 15th and 24th day after exposure—the time during which virus was detectable in the saliva of the clinical cases. All the children had developed antibody at the time symptoms appeared.

We know little of the factors which so affect the extent of infection that only in some do clinical signs and symptoms become manifest. The outcome probably depends on the extent of virus invasion up to the time antibody is developed, but it may be also that individuals vary in the extent of tissue involvement necessary for clinical symptoms to appear. We do not know whether the subclinical infection is due to partial inhibition of virus and slowing of the infection by one or other of the non-specific factors mentioned previously, or whether the speed and extent of the immunological response on the part of the host's tissues is unusually great. Many observations have shown that there may be great variation in the antibody response of individuals to the same infecting virus, a point which might have had more attention than it has received in this lecture. However, it seems likely that in most cases of subclinical as of symptomatic infection the development of antibody is mainly responsible for arresting the progress of virus invasion. There are other aspects of the subject which I have had no time to discuss, such as variation in antibody response to different viruses and to different kinds of vaccine used for active immunization, variations in the quality of antibody and variation in resistance of virus to the action of antibody. From the point of view of medical practice, however, what is known of the mode of action of antibody indicates that antisera may be successfully used to prevent illness if given at the right time, but are not likely to be of value in treatment of established disease. Limited trials of antisera in treatment have confirmed this latter expectation.

REFERENCES

[illegible]

BROWN, G. C., and FRANCIS, T. (1947). *J. Immunol.* **57**. 1.

DALLDORF, G. (1950). *The Pathogenesis and Pathology of Viral Diseases*, p 31. Columbia University Press, New York

[illegible] *J exp. Biol*

FENNER, F., and FENNER, E. M B. (1949). *Aust. J exp. Biol. med Sci* **27**, 19

FRANCIS, T, PEARSON, H. E., SALK, J E, and BROWN, P. N (1944). *Amer J publ Hlth* **34**, 317.

GINSBERG, H S, GOEBEL, W. F, and HORSFALL, F. L (1948) *J. exp. Med.* **87**, 385

HENLE, G, HENLE, W., WENDELL, K K, and ROSENBERG, P (1948) *J exp Med* **88**, 223.

MCCARTHY, K., and GERMER, W. D. (1952). (To be published.)

MARIS, E. P, ENDERS, J. F., STOKES, J, and KANE, L. W. (1946) *J. exp Med* **48**, 323

ROUS, P., MCMASTER, P. D., and HUDACK, S. S (1935) *J. exp Med* **61**, 657

SMITH, W. (1951) *J. Roy. inst. publ Hlth.* Oct., 307.

XIII

Blood Coagulation in Theory and Practice

R. G. MACFARLANE

THE easiest way of reviewing the development of blood coagulation theory is by means of a diagram illustrating the stages by which it has been built up. Even the skeleton of well-established fact is complicated and anyone who may have hoped for a simplification of his ideas will, I am afraid, be disappointed. The logical beginning for such a scheme is fibrin. So far no one has actually denied that blood does clot when withdrawn from the body, or that this solidity is due to the formation of a network of fibrin It has been recognized for nearly a hundred years that fibrin is derived from fibrinogen, one of the plasma proteins, which has a large molecule, probably of an elongated needle-like shape A fundamental point should be emphasized here. The transformation of fibrinogen to fibrin is the only indicator of coagulant activity and every factor or condition supposed to influence the clotting mechanism has to be inferred from changes in the rate of the fibrinogen-fibrin transformation. No other demonstrable change, chemical or physical, occurs in any of the clotting factors, with the possible exception of agglutination of the platelets. Fibrinogen is a relatively stable soluble protein which is changed to insoluble fibrin by a specific factor, thrombin. Certain other substances are capable of precipitating or coagulating fibrinogen but they do not form the characteristic network of natural fibrin. Thrombin is not normally present in the circulating blood but is generated during the process of coagulation by the activation of its pre-

cursor, prothrombin, a factor associated with the α-globulin fraction of the plasma. Activation of prothrombin is brought about by the action of tissue fluids or extracts in the presence of calcium and of certain other factors which I will mention later. Substances which remove ionized calcium such as citrate or oxalate prevent the activation of prothrombin. The tissue factor is usually called thromboplastin (see Fig. 1).

This is the simplest form of the original classical theory of coagulation which was built up by the work of Hewson, Hunter, Lister, Schmidt, Morawitz, Mellanby, Howell and many others. It served quite well to explain the commonly observed facts of

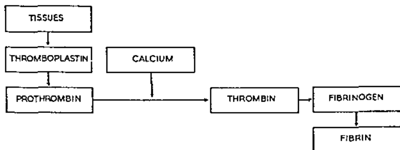


FIG. 1.

coagulation until workers began to measure the supposed concentration or activity of its various components. Quantitative experiments always put a great strain on a theory which is not complete, and in this case while the backbone of the theory has remained firm it has been found necessary to make a number of additions to the skeleton The first indications of these impending changes occurred when the one-stage method for determining prothrombin was introduced by Quick (1935) According to the classical theory if an optimum amount of calcium and excess thromboplastin is added to a mixture containing an unknown [illegible] resulted in great clinical advances, but it also revealed some anomalies which were difficult to explain. It did not give the same results for instance as the two-stage method for deter-

XIII

Blood Coagulation in Theory and Practice

R. G. MACFARLANE

THE easiest way of reviewing the development of blood coagulation theory is by means of a diagram illustrating the stages by which it has been built up. Even the skeleton of well-established fact is complicated and anyone who may have hoped for a simplification of his ideas will, I am afraid, be disappointed. The logical beginning for such a scheme is fibrin. *So far no one has actually denied that blood does clot when* withdrawn from the body, or that this solidity is due to the formation of a network of fibrin. It has been recognized for nearly a hundred years that fibrin is derived from fibrinogen, one of the plasma proteins, which has a large molecule, probably of an elongated needle-like shape. A fundamental point should be emphasized here. The transformation of fibrinogen to fibrin is the only indicator of coagulant activity and every factor or condition supposed to influence the clotting mechanism has to be inferred from changes in the rate of the fibrinogen-fibrin transformation. No other demonstrable change, chemical or physical, occurs in any of the clotting factors, with the possible exception of agglutination of the platelets. Fibrinogen is a relatively stable soluble protein which is changed to insoluble fibrin by a specific factor, thrombin. Certain other substances are capable of precipitating or coagulating fibrinogen but they do not form the characteristic network of natural fibrin. Thrombin is not normally present in the circulating blood but is generated during the process of coagulation by the activation of its pre-

As it now appears established that factor V can be regarded as a real and necessary part of the clotting mechanism it can be added to the classical theory. Its conversion to factor VI is more problematical but for completeness' sake factor VI is included in the figure though future work may modify this addition. In this particular instance the discovery of a new factor has only involved one further hypothetical factor; sometimes new factors come in batches of five, the factor itself, its inhibitor, its inactive precursor, its activator and the inhibitor of the activator!

We must now consider in greater detail what is meant by the term 'thromboplastin'. The existence of thromboplastin is inferred from an effect and no single substance can be identified as being responsible for this effect. The original view was that thromboplastic activity was a function of damaged tissue, and it is a familiar observation that almost any ground-up tissue, particularly lung, brain or kidney, will potentiate prothrombin conversion. But when blood is taken by venepuncture it clots [illegible]

the most careful technique designed to avoid contamination with tissue fluid does not prevent the blood clotting in the ordinary way. It must be assumed, therefore, that the blood itself contains its own source of thromboplastic activity and since it does not normally clot in the vessels it must be assumed further that this activity arises only when the blood is shed, or otherwise altered. It now seems probable that there are at least three factors concerned in the development of this intrinsic thromboplastic activity, these being the platelets, a plasma factor, and a 'contact factor' activated by foreign surfaces. If any one of these is missing coagulation will not occur. For instance if the blood is taken into a syringe coated with silicone, which is a non-active surface, and is then transferred to a silicone-coated vessel coagulation will not occur for many hours, though it will occur promptly if the blood is transferred to ordinary glass. If the blood collected by this silicone technique is centrifuged at high speed without contact with glass the resultant platelet-free plasma will sometimes fail to clot even

mining prothrombin in which the prothrombin is converted to thrombin in the first stage of the procedure and the concentration of thrombin achieved is then estimated by the addition of samples to fibrinogen solution in the second stage. It soon became obvious that the rate of thrombin production in the presence of thromboplastin and calcium was dependent not only upon the amount of prothrombin but on other factors hitherto unknown. The existence of one of these accessory factors was clearly revealed by the brilliant work of Owren (1947) as a result of his study in 1943 of a patient with a rare form of haemorrhagic diathesis. This patient had a clotting defect characterized by a long one-stage prothrombin time, a finding which should have implied that she had a deficiency of prothrombin This defect, however, could be corrected by the addition to her blood of an amount of normal plasma so small that the effect of the added prothrombin must have been insignificant. The defect was also corrected by a preparation of plasma from which all the prothrombin had been removed. It was soon proved that this patient lacked a normal factor which is essential for the rapid conversion of prothrombin and which is not included in the four factors of the classical theory of blood coagulation. Owren called this additional factor 'factor V' and suggested that in the process of coagulation it is converted into an active form, 'factor VI', which is a potent accelerator of prothrombin conversion. He has recently proposed the names 'accelerin' and 'pro-accelerin' for factor VI and factor V respectively (Owren, 1950). As so often happens workers in other laboratories had independently discovered the existence of factor V more or less simultaneously, though their characterization of this factor was not so clear or so complete as that of Owren It now seems probable that the Ac globulin of Ware, Guest and Seegers (1947), the labile factor of Quick (1943) and the accelerator factor of Fantl and Nance (1946) are all the same as factor V It is only fair to record that Nolf (1908, 1938) has always maintained that five factors take part in coagulation and it is probable that his thrombogen is identical with factor V. Unfortunately Nolf's work has been so obscured by a complex theory that it has not received the recognition it deserves.

hastens thrombin generation by reacting in some way with the platelets thereby stimulating the more rapid formation of thromboplastic activity, and Owren (1950) considers that thrombin is the activator of factor V. Running almost parallel with thrombin generation is thrombin destruction, at a rate which is itself proportional to the concentration of thrombin at any particular moment. It is clear that the coagulation of fibrinogen depends upon a race between thrombin generation on the one hand and thrombin destruction on the other and anything which influences either of these rates even to a relatively minor

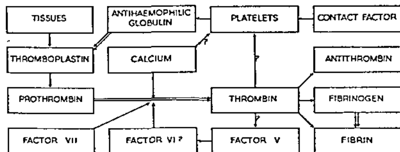


FIG 2. A scheme illustrating the possible relationship of factors concerned in blood coagulation A single arrow signifies 'reacts with', a double arrow signifies 'converted to'

degree may have a comparatively large effect on the time required for coagulation.

There are many other complications which could be mentioned in relation to this theory of coagulation but I will not dwell on them because they are not well established. There have been speculations, for instance, on the nature of thromboplastic activity, including a suggestion that it is proteolytic in nature. There have been suggestions of two stages in the thrombin-fibrinogen reaction which are supposed to be demonstrated by

blance to each other, or to what I have described, but all one can say about them at the moment is that they have little evidence to encourage their existence, and certainly not enough to challenge the foundation of the classical theory.

when in contact with glass, showing that both contact and the platelets are required for coagulation. The plasma factor is more difficult to study because like factor V it is not recognizable in normal blood, being only obvious when it is missing. It now seems likely that haemophilia is due to a partial or complete absence of this plasma factor which, since the addition of a small amount of the globulin fraction of normal plasma will correct the clotting defect in haemophilia, has come to be called 'antihaemophilic globulin'. Haemophilic plasma even without contact and without platelets can be clotted readily by the addition of tissue extracts. It can also be clotted, in the presence of platelets and contact with a foreign surface, by the addition of a small amount of antihaemophilic globulin. The absence of any one of these three factors inhibits its coagulation. It is therefore *reasonable to suppose that these three factors together result in* the development of a thromboplastic activity which is similar to the activity of tissue extract. The co-operation of these three factors can therefore be illustrated in the diagram (Fig. 2) as being probably responsible for the development of thromboplastic activity in the blood, the tissues providing an alternative source of such activity

Opposing these coagulant factors there are a number of inhibitors which play an important part in limiting coagulation to its normal function. The most intensively studied of these are the antithrombins which together remove thrombin with a rapidity almost equal to the rate of its formation One of these antithrombins is associated with the albumin fraction of the plasma and is greatly potentiated by heparin, and thrombin is strongly adsorbed by fibrin which is responsible for a considerable part of the antithrombic action of clotted blood. There is probably also an antithromboplastin though its nature and activity have not been clearly defined. Tocantins (1945) believes that this inhibitor may be present in excess in haemophilia.

In operation the coagulation system behaves in a dynamic manner which is extremely complicated. The rate of thrombin generation once it begins proceeds with increasing velocity suggesting the possibility of some autocatalytic effect. To explain this it has been supposed by Quick (1951) that thrombin itself

inferred from an effect produced by a number of factors probably working in co-operation. Yet throughout the literature the term has been used as if something chemically specific is being described. Sometimes workers who use the term thromboplastin, or thrombokinase, define what they mean by it, and then it is found that various extracts of brain, lung, kidney, placenta, human milk, or Russell viper venom with or without lecithin, are all called 'thromboplastin'. Some workers even go further using the word 'cephalin' instead of thromboplastin. Cephalin is a definite chemical entity which was thought some years ago to be the active principle in thromboplastin, a view that is now generally discarded. Yet Tocantins (1945) refers to cephalin when he actually uses brain extract in his experiments, and supposes that an 'anticephalin' is the cause of the clotting defect in haemophilia when he observes some substance capable of inhibiting general thromboplastic activity.

The real factors in blood coagulation have suffered similar abuse. The term 'prothrombin' may mean simply a crude fraction of the plasma containing many other things active in coagulation. 'Fibrinogen' may be a relatively pure protein prepared with great care, or it may actually be oxalated plasma, which is of course complete with most of the clotting factors known to man A sad and important example of this misuse of words is provided by the work of Addis (1911). In the process of investigating the problem of haemophilia he proceeded to examine in a most methodical way each of the clotting factors in this condition then known to exist, his work constituting a classical contribution to this subject. In order to investigate the activity of prothrombin in haemophilia he made 'prothrombin' preparations from normal and haemophilic plasma by precipitating the diluted plasma with CO_2, dissolving the precipitates in saline, and removing fibrinogen by the addition of a small amount of thrombin He found that a minute proportion of the solution derived from normal plasma would correct the clotting defect of haemophilic blood whereas the corresponding preparation from haemophilic plasma had no such effect. He concluded from this that there was a qualitative defect in haemophilic prothrombin, and for many years afterwards other workers

SOME PITFALLS IN BLOOD COAGULATION RESEARCH

Most of the work that has gone into the scheme outlined in the previous section has been based on sound experiment. There has been a good deal of repetition, partly because many workers seem to be unfamiliar with the published work of others, partly because they have used different terminologies so that essentially the same factors and the same experiments are described again and again in different terms. Table 1 gives a list of some of the

TABLE 1. Synonyms for the Coagulation Factors

Prothrombin	*Factor V* (Owren)
Thrombogen (Morawitz) Thrombozyme (Nolf) Proserozyme } (Bordet) Serozyme } Inactive Prothrombin } (Quick) Active Prothrombin }	Thrombogen (Nolf) Labile Factor (Quick) A.C. Globulin (Seegers) Prothrombin Accelerator (Fantl and Nance) ?Prothrombokinase (Milstone)
Factor VII (Koller)	*Thromboplastin* (Howell)
Convertin (Owren) ? Co. thromboplastin (Magath and Hurn) S.P.C.A. (Alexander and de Vries) ? Serum A.C. Globulin (Seegers) P.C.A. (Jacox)	Thrombokinase (Morawitz) Cytozyme (Bordet) Cephalin (Tocantins) Kinase (Mellanby) Fibrinogen A (Wooldridge) Thrombokinin (Lengenhager)

synonyms for the more important clotting factors. There are many examples, however, of practices and reasoning which have led to endless confusion, and it is a few examples of these which I propose to describe in the following section

A great deal of trouble has arisen from the very common use of a specific term to describe not a single substance, but perhaps a group of substances or even some effect or phenomenon, a tendency by no means limited to blood coagulation work. An example of this is the use of the term 'thromboplastin' Thromboplastin is not an actual substance, it is a hypothetical material

tively little effect. It was argued from this that the distilled water lysed the haemophilic platelets thus releasing their supposed content of prothrombin, and that the platelets were unduly stable in haemophilia since their lysis by distilled water corrected the clotting defect. It is now known that dilution of haemophilic plasma with saline causes a similar shortening of the clotting time irrespective of platelet lysis, and had this effect of dilution been controlled the argument based on the effect of distilled water would have been seen to be incorrect.

Another source of confusion is the building of tentative hypotheses and their subsequent transformation in the literature from conjecture into statements of fact. Brinkhous (1947) and Quick (1947) independently carried out an almost identical experiment, and obtained almost exactly the same results. They drew quite different conclusions. The experiment consisted in preparing normal and haemophilic platelet-free plasma by using silicone-coated apparatus, and observing the effects of the addition of both normal and haemophilic platelets to either type of plasma. It was found that either normal or haemophilic platelets were equally capable of clotting the normal platelet-free plasma in glass vessels but that neither normal nor haemophilic platelets were capable of clotting the haemophilic plasma unless a small amount of normal plasma was added at the same time. This fundamental experiment demonstrated conclusively that haemophilic platelets behaved in the same way as normal platelets, and that haemophilic plasma requires something present in normal plasma before coagulation, in the presence of platelets and contact with a foreign surface, can take place. Two conflicting hypotheses as to the mode of interaction of the three factors were then put forward by these workers. Quick suggested that the plasma factor was a precursor of thromboplastin which he called 'thromboplastinogen'. This thromboplastinogen was in his view activated by some factor released by the platelets which disrupted on contact with a foreign surface. Brinkhous on the other hand supposed that the platelets contained thromboplastin and that they were disrupted by a plasma factor called by him 'thrombocytolysin' which was activated by contact with glass. Both workers agreed that the plasma factor was deficient

attempted to prove or disprove the existence of this supposed prothrombin defect. Some twenty-five years later Patek and Taylor (1937) and Bendien and Van Creveld (1937) showed that a small proportion of a 'globulin substance' derived from normal plasma was capable of shortening the clotting time of haemophilic blood. The globulin substance was obtained by precipitating normal plasma with acid and similar preparations from haemophilic plasma were inactive. This work was generally recognized as a great advance in the research on haemophilia, and it led directly to the concept of antihaemophilic globulin and to the preparation of material so active that as little as 10 or 20 mgm given by injection will improve the clotting defect in a haemophilic subject over a period of some hours. [illegible]

pointed out its extraordinary activity the situation would have been perfectly clear. Since he called it prothrombin it was assumed by other workers that he meant prothrombin and when later the real prothrombin was shown to be normal in haemophilia Addis' work was largely disregarded. In this instance the proper use of words might have advanced research in haemophilia by twenty-five years.

A familiar cause of misinterpretation arises from a failure to carry out the proper controls. For instance it was supposed for many years that the cause of the clotting defect in haemophilia was an increased stability of the platelets. This was based on the observation that the addition of normal platelets to haemophilic blood markedly shortened the clotting time. It was not until Patek and Stetson (1936) found that both normal and haemophilic platelets were equally efficient in clotting platelet-free normal plasma and that it was the small amount of normal plasma contaminating the normal platelets which was responsible for the shortening of the clotting time of haemophilic blood that the fallacy was revealed. The confusion meanwhile had been further increased by the observation that the addition of distilled water to haemophilic plasma shortened its clotting time, whereas a similar addition to normal plasma had rela-

I hope that these examples are sufficient to emphasize how important it is not to go beyond the facts in any experiment and to make it clear exactly what has been done. It is a crime to talk about complement when one means guinea-pig serum, or cephalin when one means brain extract, or to make up a complicated theory, if by so doing one obscures a good experiment. Some of the experiments which have been hailed with acclaim during the last ten years were carried out in some form by Nolf [illegible]

folding, but so often the constructors of scaffolding of this sort become so enamoured of it that they refuse to allow permanent building to proceed in case it might be altered or obscured.

BLOOD COAGULATION IN PRACTICE

The question which the ordinary pathologist and clinician asks of any theoretical work is, 'How does it apply to practical problems?' In blood coagulation the answer is that practical advances of the greatest importance have been directly contributed by the application of theory to clinical investigation. One has only to cite the recognition and treatment of Vitamin K deficiency, the discovery and use of heparin, dicoumarin and tromexan, and the use of silicones, ion exchange resins and such commonplace things as citrate and oxalate to realize how large this contribution has been. An important advance too is the introduction of a number of tests of clotting function which are valuable in the differential diagnosis of a variety of haemorrhagic states. The oldest and simplest of these tests is the whole blood coagulation time, which is best carried out by adding measured amounts of blood to small test-tubes kept at 37° and observing the time required for coagulation As so often happens with simple tests the interpretation of the results is most complicated because they depend on the complete range of clotting factors and inhibitors. If a patient has a long clotting time there is clearly something wrong with his coagulation mechanism, but there is no indication as to which part of the mechanism may be at fault But, on the other hand, a normal coagulation time

in haemophilia. Such hypotheses are of course perfectly permissible, but only as hypotheses, since the available evidence does not prove the correctness of either one or the other and many explanations of the observed facts could be invented. But it is now stated categorically by Quick (1951) that haemophilia is caused by 'hypothromboplastinogenaemia' as if this were an established fact and he refers to the experiment described above as furnishing the proof of this statement.

The most recent work of Owren (1950) has also led to a certain amount of confusion, in this instance relating to the identities of the factors being studied. Owren observed that a mixture of human serum and Seitz-filtered ox plasma generates thrombin Seitz filtration is supposed to remove prothrombin, so that if the ox plasma contains no prothrombin the thrombin generation must have been provided by unsuspected prothrombin contained in the human serum. Since this serum did not generate thrombin on the addition of thromboplastin Owren concluded that some factor required for prothrombin activation, other than factor V and thromboplastin, was missing in the human serum and was provided by the Seitz-filtered ox plasma. He called this factor 'convertin' and suggested that it is the direct activator of prothrombin, being derived from an inactive precursor called 'pro-convertin' which is itself activated by thromboplastin. It was supposed that convertin is consumed during coagulation before prothrombin conversion is complete, so that the serum contains a proportion of unconverted prothrombin which will not react with thromboplastin since no convertin is available. Koller, Loeliger and Duckert (1951) have repeated these experiments and though they obtained much the same results, they have come to a diametrically opposed conclusion. They have shown that it is the Seitz-filtered ox plasma which contains prothrombin, and that it lacks a conversion factor which is presumably removed by Seitz filtration, but which is present in the human serum. This factor they have called 'factor VII' and they consider that it is, like factor V, an accelerator of prothrombin conversion It appears from his paper that Owren has conceded this point so that presumably convertin and pro-convertin as described by him on the basis of his original experiment do not exist.

defect of known haemophilic plasma then it presumably has a normal complement of antihaemophilic globulin. If it is less effective it probably has a deficiency of antihaemophilic globulin.

TABLE 2

Test	Finding	Possible Defect
Whole blood Clotting time	Long	Any known factor. Presence of inhibitor.
	Normal	Any known factor partially deficient.
1 stage 'Prothrombin Test'	Long	Fibrinogen. Prothrombin. Factors V or VII Presence of antithrombin.
	Normal	Antihaemophilic globulin Platelets Antithromboplastin present.
2 stage 'Prothrombin Test'	Abnormal	Prothrombin. (Test less susceptible to deficiency of accelerators than 1 stage)
Prothrombin Consumption Test	Abnormal	Platelet Antihaemophilic factor Antithromboplastin present
Addition to Normal Plasma	Clotting time prolonged	Presence of anticoagulant.
Addition to Haemophilic Plasma	Clotting time normal	Not A H G. defect
	Clotting time long	Probably A H G defect.

In many cases a clotting defect can be shown to be due to the presence of inhibitors rather than to a deficiency of clotting factors. The presence or absence of inhibitors can be demonstrated by adding a proportion of the patient's plasma to normal plasma and observing whether or not the clotting time of the

does not mean that the mechanism is normal. Quite gross deficiencies of prothrombin, fibrinogen, platelets or antihaemophilic globulin may occur without much alteration in the time required for the appearance of the first strands of fibrin.

The next test in order of complexity is the one-stage 'prothrombin time' devised by Quick, which consists of adding brain extract and calcium to oxalated plasma and observing the time of coagulation. This was supposed to be a direct measure of prothrombin concentration, and even though this contention is not valid, it has proved to be a most useful test, since it reveals the clotting defect of vitamin K deficiency and controls very well the use of anticoagulant therapy. If the 'prothrombin time' by this method is long there may be a deficiency of prothrombin or of the accelerators of prothrombin activation, a shortage of fibrinogen or an excess of antithrombin. If it is normal despite a long whole blood clotting time then a deficiency of the intrinsic thromboplastin mechanism is likely to be present. If it is necessary to obtain a more precise estimate of prothrombin itself, then one must carry out a two-stage prothrombin estimation which is a measure of the amount of thrombin produced rather than of the rate of thrombin production. The two-stage method and particularly its more recent modifications are less sensitive to a deficiency of accelerators than the one-stage method.

A fairly recent development is the 'prothrombin consumption test'. In this the amount of prothrombin remaining in the serum after coagulation is measured by means of the addition of thromboplastin and estimation of the thrombin produced. Normally there is relatively little prothrombin remaining after coagulation is complete, but if there is a deficiency of intrinsic thromboplastin then the serum may contain large amounts of prothrombin which has not been converted. Abnormal prothrombin consumption, as shown by an excess remaining in the serum, occurs if there is a deficiency of the platelets, or antihaemophilic globulin deficiency (Merskey, 1950a), or an excess of antithromboplastin. A deficiency of antihaemophilic globulin may be detected by observing the effect of mixing the suspected plasma with known haemophilic plasma. If the suspected plasma is as capable as normal plasma of correcting the clotting

Haemophilia is probably the most important of the haemorrhagic diseases because it is hereditary and causes a life-long and most serious disability. It is a sex-linked condition affecting males, the heterozygous female being apparently normal. In the

TABLE 3

Condition	Defect	Causation
Fibrinopenia	Absence or deficiency of Fibrinogen	Inherited.
		Liver damage.
		Fibrinolysis
Heparinaemia	Increase of Antithrombin	Anaphylactic shock Irradiation.
Haemorrhagic Jaundice	Hypoprothrombinaemia	Vit. K deficiency.
Haemorrhagic Disease of Newborn		
Idiopathic Hypoprothrombinaemia		Unknown
Para-haemophilia	Factor V deficiency	Unknown.
Pseudo-haemophilia	Antithromboplastin deficiency or destruction of anti-haemophilic globulin	Immunization Idiopathic
Haemophilia	Deficiency of anti-haemophilic globulin	Sex-linked inheritance.

extremely rare event of a homozygous female being produced as the result of the marriage of a haemophilic male with a heterozygous female carrier of haemophilia, such homozygous females have all the clinical manifestations of haemophilia (Merskey, 1951; Israels, Lempert and Gilbertson, 1951). Haemophilia is almost certainly due to a deficiency of intrinsic thromboplastin

latter is prolonged. If prolongation occurs further investigations are required to identify the inhibitor. It may be of the heparin type, in which case the inhibitory effect is abolished by adding protamine or toluidine blue. It may be an antithromboplastin, when incubation of the patient's plasma with brain extract causes a reduction of the latter's thromboplastic activity. It may be an inhibitor of antihaemophilic globulin, as shown by the fact that after mixture with the patient's plasma the normal plasma loses its power to correct the clotting defect of known haemophilic blood.

The clinical conditions in which these defects occur are given in Table 3. Fibrinogen deficiency is a very rare cause of abnormal bleeding; when it occurs it may be as the result of an inherited defect, or as the result of severe liver damage, or may be occasionally due to excess *fibrinolysis. In* such cases the blood either does not clot at all or if it does very friable and incomplete clots are produced. Prothrombin deficiency is a much more important condition occurring in any state in which the intake of vitamin K is impaired, such as obstructive jaundice, various forms of sprue and in the newborn. Prothrombin deficiency may occasionally occur idiopathically. It is now frequently produced artificially by the administration of such drugs as dicoumarin and tromexan which apparently interfere with prothrombin production, but which also probably cause a reduction of one or more of the accelerators of prothrombin conversion. It is necessary to control such anticoagulant therapy by estimation of the clotting defect produced, and this is most conveniently done by the 'one-stage prothrombin time' method. Unfortunately a great deal of confusion is caused by the fact that different units are used by different laboratories in which the results of this test are reported. The 'prothrombin index' for instance though expressed as a percentage gives a picture of the severity of the artificially produced coagulation defect entirely different from the supposed 'prothrombin concentration' which is read from a calibration curve and is also expressed as a percentage. It is essential to be absolutely certain what units are being used, and what their significance is, before the results of these 'prothrombin determinations' are considered (Biggs, 1951).

antibody arising by a mechanism similar to that responsible for erythroblastosis foetalis. It is significant that a number of genuine haemophilics have developed an antibody of this type after repeated blood transfusions. Such immunization accounts for the fact that the normal beneficial effects of blood transfusion or administration of antihaemophilic globulin have been lost in certain patients. This tendency to immunization against antihaemophilic globulin is a very serious stumbling block in the treatment of haemophilia, and it can only be hoped that future work will provide a non-antigenic form of antihaemophilic material active enough to be given in minute doses comparable with those of insulin or vitamin B12.

As a sort of recapitulation I would like to give the results of a new test which Dr. Rosemary Biggs has been studying during the past few months in Oxford. This test consists of measuring the amount of thrombin which is present before, during and after coagulation. Undiluted whole blood, or recalcified plasma, is allowed to clot in glass tubes and small samples are taken at intervals for thrombin assay. The effect of the addition of thromboplastin or thrombin, the removal or addition of platelets and haemophilic globulin, the effect of silicone surfaces and so on can be examined in this way under conditions which are nearer to natural clotting than any other analytical coagulation test From the thrombin measurements obtained a curve of thrombin concentration is constructed. The normal curve is shown as curve 1 in Fig 3. It can be seen that after a lag phase of about three minutes, during which no thrombin is produced, a sudden generation of thrombin occurs so the concentration rises steeply to a maximum, and then almost as rapidly declines. The fall in thrombin concentration is of course due to its destruction by antithrombin Curve 2 in Fig 3 is obtained if a very small amount of thrombin is added to the blood immediately it is transferred to the test-tube. The effect of this addition is to shorten the lag phase so that thrombin generation begins about two minutes earlier, but the rate of thrombin generation is not significantly increased. This observation emphasizes the fact that thrombin plays some part in thrombin generation. Fig. 4 shows thrombin generation curves obtained from recalcified

and this in turn is probably due to a shortage of the plasma factor or antihaemophilic globulin. When the clotting defect is well marked and the family history is typical diagnosis is easy, but in some cases the whole blood clotting time is relatively normal and this fact coupled with a possible absence of family history makes it necessary to confirm the diagnosis by other tests. Of these the prothrombin consumption test and measurement of the patient's antihaemophilic globulin activity against known haemophilic blood have proved to be the most reliable (Merskey, 1950b). Unfortunately it is not possible to detect any clotting abnormality in the female carriers of the haemophilic gene, though we hope that this may be achieved in the future (Merskey and Macfarlane, 1951). If it could be determined with certainty that a particular woman was likely to pass the condition on to her children it is probable that she would elect to remain childless and the incidence of haemophilia would be greatly reduced in a very few generations.

It is now becoming recognized that an increasing number of clinical states may be caused by the presence of naturally occurring anticoagulants. For instance a heparin-like substance may occur in certain conditions such as anaphylactic shock, after exposı
taneously.
clotting time
the anticoagulant can be demonstrated by the fact that this inhibitory effect is neutralized by the addition of toluidine blue or protamine. A number of cases of haemophilia-like disease, or pseudo-haemophilia, clinically almost indistinguishable from true haemophilia but acquired in adult life by both men and women, have been reported. This condition has in most cases been shown to be due to the presence of an inhibitor which seems to act against thromboplastin or against antihaemophilic globulin. The presence of such anticoagulants will create a

number of such cases have occurred shortly after pregnancy and one wonders whether this anticoagulant may be some form of

antibody arising by a mechanism similar to that responsible for erythroblastosis foetalis. It is significant that a number of genuine haemophilics have developed an antibody of this type after repeated blood transfusions. Such immunization accounts for the fact that the normal beneficial effects of blood transfusion or administration of antihaemophilic globulin have been lost in certain patients. This tendency to immunization against antihaemophilic globulin is a very serious stumbling block in the treatment of haemophilia, and it can only be hoped that future work will provide a non-antigenic form of antihaemophilic material active enough to be given in minute doses comparable with those of insulin or vitamin B12.

As a sort of recapitulation I would like to give the results of a new test which Dr. Rosemary Biggs has been studying during the past few months in Oxford. This test consists of measuring the amount of thrombin which is present before, during and after coagulation. Undiluted whole blood, or recalcified plasma, is allowed to clot in glass tubes and small samples are taken at intervals for thrombin assay. The effect of the addition of thromboplastin or thrombin, the removal or addition of platelets and haemophilic globulin, the effect of silicone surfaces and so on can be examined in this way under conditions which are nearer to natural clotting than any other analytical coagulation test. From the thrombin measurements obtained a curve of thrombin concentration is constructed. The normal curve is shown as curve 1 in Fig. 3 It can be seen that after a lag phase of about three minutes, during which no thrombin is produced, a sudden generation of thrombin occurs so the concentration rises steeply to a maximum, and then almost as rapidly declines. The fall in thrombin concentration is of course due to its destruction by antithrombin. Curve 2 in Fig. 3 is obtained if a very small amount of thrombin is added to the blood immediately it is transferred to the test-tube. The effect of this addition is to shorten the lag phase so that thrombin generation begins about two minutes earlier, but the rate of thrombin generation is not significantly increased. This observation emphasizes the fact that thrombin plays some part in thrombin generation Fig. 4 shows thrombin generation curves obtained from recalcified

plasma. Curve 1 is obtained from simple recalcification without the addition of thromboplastin. Curve 2 is from a similar plasma sample to which enough brain extract has been added to cause coagulation within the first half-minute. It will be observed that

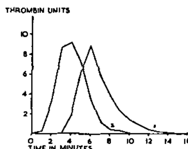


Fig. 3 Thrombin generation curves from normal whole blood. Curve 1, whole blood alone; curve 2, blood to which a small amount of thrombin has been added at zero time.

the addition of brain extract shortens the lag phase of thrombin generation but it does not materially increase the speed at which thrombin is generated. From these two experiments it can be deduced that in whole blood or recalcified citrated plasma in

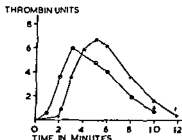


Fig 4. Thrombin generation curves from recalcified citrated plasma. Curve 1, recalcified plasma alone, curve 2, recalcified plasma to which brain extract has been added at zero time.

contact with glass no thrombin is generated, and hence no thromboplastic activity is present until just before coagulation occurs, that is for several minutes after withdrawal of the blood or recalcification. Once thrombin generation begins it proceeds with a rapidity which is as great as that observed when sufficient brain extract has been added to cause coagulation in twenty to

thirty seconds. It appears therefore that the intrinsic thromboplastin has a potential activity as great as that of the brain extract commonly used for the one-stage prothrombin test. Since the clotting time of whole blood without added thromboplastin may be six minutes or more it has always been supposed that the intrinsic thromboplastin is relatively very feeble since this clotting time is compared with the fifteen or twenty seconds of the one-stage prothrombin test in which thromboplastin is added. It can now be seen that this view is probably wrong and that the

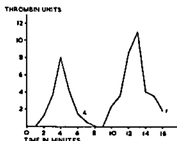


Fig. 5 Thrombin generation curves from whole blood allowed to coagulate in partially siliconed tubes Curve 1, blood alone, curve 2, blood to which a small amount of thrombin has been added at zero time

intrinsic thromboplastic activity, which does not appear until just before normal clotting, is far greater than has been supposed hitherto.

The time at which intrinsic thromboplastic activity appears depends upon the degree of contact with a foreign surface. Fig. 5, curve 1, shows the effect of carrying out thrombin generation in partially siliconed apparatus. There is an increase in the lag phase although the rate of thrombin generation when it occurs is normal This lag phase can be greatly reduced by adding thrombin at the beginning of the test (curve 2, Fig. 5) suggesting that the accelerating action of thrombin on thrombin formation is independent of contact. The effect of platelets is illustrated in Fig. 6 which shows a series of thrombin generation curves obtained from recalcified plasma samples in which the platelet count has been successively reduced by centrifuging. There is not a great change in the time at which thrombin begins to be generated but there is a great reduction in the

amount of thrombin generated in the samples with reduced platelet counts. This suggests that the platelets are quantitatively related to the amount of thromboplastic activity which appears, but not to the time of its appearance.

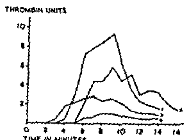


Fig 6. Thrombin generation curves from samples of recalcified plasma containing progressively fewer platelets from curve 1 (normal platelet count) to curve 4 (1 per cent of normal count).

In haemophilia characteristic thrombin generation curves are obtained, and the abnormality is well marked even in those cases in which the whole blood coagulation time falls within normal limits. In Fig. 7 curves from mixtures of haemophilic

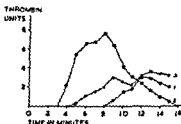


Fig 7 Thrombin generation curves from mixtures of haemophilic and normal plasma Curve 1, 99 per cent haemophilic plasma, 1 per cent normal, curve 2, 90 per cent haemophilic plasma, 10 per cent normal, curve 3, 80 per cent haemophilic plasma, 20 per cent normal.

plasma and normal plasma are illustrated The thrombin generation curve of haemophilic plasma alone would be off the page to the right, and curve 1 is derived from haemophilic plasma with 1 per cent normal plasma, having a clotting time of about nine minutes Curve 2 is from haemophilic plasma with 10 per cent normal plasma with a clotting time of five minutes, but still a considerable diminution in the speed of thrombin

generation. Curve 3 is from a mixture of 20 per cent normal plasma and 80 per cent haemophilic plasma and at this level of added antihaemophilic globulin the thrombin generation curve is normal. These observations illustrate that a reduction of the whole blood clotting time to normal is by no means a proof that the clotting function is restored to normal. The thrombin generation test appears to be a very delicate indicator of overall clotting efficiency.

In conclusion it is not necessary to say any more about blood coagulation in practice. The advances in diagnosis and treatment which I have already mentioned are more than capable of speaking for themselves. But what about the theory? Most people would say that it is far too complicated, they cannot believe that nature really needs all these factors, co-factors, accelerators, activators and inhibitors just to clot fibrinogen. They suggest that part at least of this complication must be due to artifacts, that we have produced many of these supposed factors in the course of our experiments and that they do not exist in real life. Until fairly recently I think that I would have agreed with this view, but now I am beginning to believe that all these factors probably do exist, because in clinical practice living examples can be found of haemorrhagic diseases caused by the absence of any one of the factors I have mentioned or by the presence of one of the inhibitors. I believe, in fact, that the theory I have described is only a small part of the actual mechanism and that many other factors are still to be discovered. But why, you will ask, is this particular biological process of fibrin formation so exceptionally complicated? The answer probably is that it only appears to be exceptionally complicated because it has been investigated to a rather exceptional degree. I think that one must believe that all biological processes, however apparently simple or unimportant, will prove to be almost infinitely complicated, and that nature will always be ahead of the research worker.

amount of thrombin generated in the samples with reduced platelet counts. This suggests that the platelets are quantitatively related to the amount of thromboplastic activity which appears, but not to the time of its appearance.

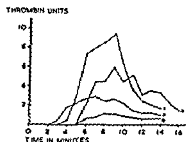


FIG. 6. Thrombin generation curves from samples of recalcified plasma containing progressively fewer platelets from curve 1 (normal platelet count) to curve 4 (1 per cent of normal count).

In haemophilia characteristic thrombin generation curves are obtained, and the abnormality is well marked even in those cases in which the whole blood coagulation time falls within normal limits. In Fig. 7 curves from mixtures of haemophilic

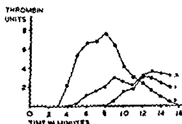


FIG 7 Thrombin generation curves from mixtures of haemophilic and normal plasma Curve 1, 99 per cent haemophilic plasma, 1 per cent normal, curve 2, 90 per cent haemophilic plasma, 10 per cent normal; curve 3, 80 per cent haemophilic plasma, 20 per cent normal.

plasma and normal plasma are illustrated. The thrombin generation curve of haemophilic plasma alone would be off the page to the right, and curve 1 is derived from haemophilic plasma with 1 per cent normal plasma, having a clotting time of about nine minutes. Curve 2 is from haemophilic plasma with 10 per cent normal plasma with a clotting time of five minutes, but still a considerable diminution in the speed of thrombin

XIV

The Contribution of Science to the Practice of Health in the First Quarter of the Twentieth Century

J. M. MACKINTOSH

In the first year of the twentieth century the fourth edition of Osler's *Principles and Practice of Medicine* was published. This book had the advantage over many text-books of having been distilled through a single mind—and a broad, humane mind at that. From this vantage point we can look back to what Sir William Osler felt to be the really fundamental changes of the previous decade and forward to a half-century's development of the practice of health. At this point, however, I must set certain boundaries to my subject. I am not concerned with the effects on the practice of health of administrative provisions under statute. These created such services as infant welfare, school health work, protection of infant life, insurance against sickness and unemployment, and pensions for widows and the aged. Their influence on the practice of health was very great indeed; but they are described in the chronicles of public health, and central and local government. Nor am I concerned primarily with discoveries, invaluable as they have been, which provided an effective remedy for certain disorders. These include the sulphonamides, penicillin, and the more recent antibiotics, as well as many of the great advances in surgery and its branches. I have limited the scope of my subject even more closely to developments in science which have influenced the practice of health through their widespread application—dreams that came true and questions which found an answer.

REFERENCES

ADDIS, T. (1911). *J. Path. Bact.* **15**, 427.
BENDIEN, W. M., and VAN CREVELD, S. (1937). *Amer. J. Dis. Child.* **54**, 713
BIGGS, R. (1951). *Prothrombin Deficiency.*
BRINKHOUS, K. M. (1947). *Proc. Soc. exp. Biol., N.Y.* **66**, 117.
FANTL, P., and NANCE, M. (1946). *Nature, Lond.* **158**, 708.
ISRAELS, M. C. G., LEMPERT, H., and GILBERTSON, P. (1951). *Lancet*, **1**, 1375
KOLLER, F., LOELIGER, A., and DUCKERT, F. (1951). *Acta haemat* **6**, 1.
LYONS, R. N. (1945). *Aust. J exp. Biol. med. Sci.* **23**, 131.
MERSKEY, C (1950a). *J. clin. Path.* **3**, 130.
— (1950b). *J. clin Path*, **3**, 301.
— (1951). *Quart. J. Med.* **20**, 79.
— and MACFARLANE, R. G. (1951). *Lancet*, March, 487.
NOLF, P. (1908) *Arch internat Physiol.* **6**, 1.
— (1938). *Medicine*, **17**, 381
OWREN, P. A. (1947). *Acta med. scand.* Suppl. 194.
— (1950). *Proc Internat. Soc. Haemat. 3rd Congress* (Greene and Stratton, New York), p. 379.
PATEK, A J, and STETSON, R. P. (1936). *J. clin Invest.* **15**, 531.
— and TAYLOR, F. H. L (1937). *J clin Invest.* **16**, 113
QUICK, A J (1935). *J biol. Chem.* **109**, lxxiii.
— (1943). *Amer J. Physiol* **140**, 212.
— (1947). *Amer. J. med Sci* **214**, 272.
— (1951) *The Physiology and Pathology of Hemostasis.*
TOCANTINS, L. M (1945). *Am. J Physiol* **143**, 67.
WARE, A. G., GUEST, M M, and SEEGERS, W. H. (1947). *J. biol. Chem.* **169**, 231

Reverting now to the clinical background, we find Osler much more at his ease, and in many respects showing insight beyond his period. His wise appreciation of appendicitis is a case in point:

'Recovery is the rule', he observes candidly. 'Out of 264 cases at St. Thomas's Hospital with the above-mentioned clinical characters [i e. the acute disease] 190 recovered ... There are surgeons who claim that the getting well in these cases does not mean much; that the patients have recurrences and are constantly liable to the graver accidents of the disease. This, I feel sure, is an unduly dark picture.'

So far, he upholds tradition; but in discussing prognosis he leads us a step further by saying, 'It is the element of *uncertainty* in individual cases which has given such an impetus to the surgical treatment of the disease'. And finally, when he comes to treatment he quietly throws in his hand·

So impressed am I by the fact that we physicians lose lives by tem-

The next matter of interest in Osler's work, and in medical text-books generally in the early years of our century, is the enormous importance attached to typhoid fever as the most common of all continued fevers and one of the most fatal His collected death-rates vary from 7 to 20 per cent, with the ill-boding record of 21 among the troops in South Africa, up to March 1901. Like other good physicians, Osler stresses the difficulties of diagnosis, in spite of the fact that the Widal reaction and direct blood culture were already widely operative. 'Errors in diagnosis are inevitable, even under the most favourable conditions. Let the "cock-sure" physician, who never makes mistakes, read the Report of the U.S. Army Commission on Typhoid during the Spanish-American War' He is already armed with figures about preventive inoculation (i e. in March 1901); and refers to Wright, of Netley—our own Almroth Wright, one of the great investigators of that era who have contributed so much to

Sir William Osler was a great humanist, and his text-book contains many apt allusions to history, and vivid phrases which have become part of our medical heritage, such as his comment on the prevention of venereal disease: 'Certainly it is better, as St Paul says, to marry than to burn, but if the former is not feasible there are other altars than those of Venus upon which a young man may light fires.'

Nevertheless, there are two features of his book which strike us today as curiously lacking in perception; and even for their time—for they persist in later editions—they were not in the van of medical thinking. The first of these may seem rather technical at first sight: it is the division and subdivision of medicine into systems, names, and fragments of the body. A careful study of the work reveals remarkably few references to the patient as a whole organism—still less as a person. So striking is the omission, that references such as the following stand out prominently, and this is itself a quotation: 'As Sir William Roberts well says, "nowhere perhaps is it more necessary than in gout to consider *the man as well as the ailment, and very often more the man* than the ailment".'

The second feature is Osler's lack of a deep understanding of the social conditions under which so many of his patients lived. It is perfectly true that even in the 1901 edition there are many references to environment—but they are perfunctory and curiously remote. Indeed, a comment on the wretched surroundings in which a disease is hatched is not infrequently followed by a recommendation that the patient should have a holiday in a warm, sunny climate, with plenty of butter and eggs and milk. Or again, his breezy remark about peptic ulcer has a chill about it: 'Servant girls', he says, 'seem particularly prone to the disease. This is to be explained partly by their careless habits in eating, partly *in connection with the associated anaemia.*' It is fair to add that this reference was omitted in later editions. You do not look to Baltimore, still less to London, for a genuine appreciation of the social factor in medicine, during the first decade of this century. You turn your wondering eyes to Boston, Mass., and to Edinburgh—to Richard Cabot in Boston, and to Robert Philip, William Ballantyne and John Thomson in Edinburgh.

to the discovery by Grijns that a diet containing the husks of rice prevented the disease. This discovery led the way to a dietetic, rather than a medical, approach to a number of other nutritional disorders; and it was the change of outlook, the illumination of the mind, rather than the specific discovery that contributed to the practice of health.

The disease called 'gout' is described frankly as a nutritional disorder; but the description in the text suggests that this means 'more fuel in the form of meat and drink than the machine needs'. Osler, adds, however, that in England the combination of poor food, defective hygiene, and an excessive consumption of malt liquors makes 'poor man's gout' a common affection.

Diabetes also is described as a nutritional (or perhaps 'metabolic' would be the better word) disorder. The work of Opie, in W. H. Welch's laboratory, showing the function and importance of the islets of Langerhans is naturally given prominence. By the *beginning of the century the internal secretion was recognized*, and called 'insulin' by Schafer in 1913. It was not until 1921, however, that the substance was isolated. The work of Banting and Best from 1921 onwards is one of the classics of twentieth-century research. Quite apart from its value to the diabetic, it has made a great addition to our knowledge of metabolism, and to the practice of living.

Perhaps the most striking of the conquests that could be consolidated on a national scale is rickets. In the 1901 edition Osler states that in London from 50 to 80 per cent of all the children we saw in schools had gross manifestations of rickets. He gives an excellent description of the disorder and indicates that lack of sunlight and fresh air are the chief predisposing causes. He mentions Bland-Sutton's experiment with lion-cubs at the Zoo, when by adding milk, pounded bones and cod-liver oil to the meat diet, rickets was prevented and the cubs were reared—for the first time in the history of the Society. With this information, it is interesting to note how Osler falters in prescribing treatment, and preventive measures are hardly mentioned, except by implication His reference to cod-liver oil, among a large array of exhibitions, is half-hearted. This is not surprising, because it was not until the great work on accessory food substances was pub-

the prevention of disease. Osler's preliminary figures show that out of 1,705 persons inoculated at Ladysmith, only 2 per cent were attacked, whereas out of 10,529 not inoculated, 14 per cent were attacked. *Among the inoculated the mortality was 0·46* per cent, and of the rest, 3 per cent.

In some instances we can see how near the good observer comes to the truth—and yet how dense is the curtain of mystery. Take at random the description of typhus:

[illegible]

time. . . .

We need not pause over lobar pneumonia, 'the most fatal of all acute diseases, killing more than diphtheria, and outranking even tuberculosis as a cause of death'—with a hospital mortality of 20 per cent. Some of our elder family doctors, however, might cast a wistful backward glance, remembering how gentle pneumonia used to be with the aged.

Our study must be *focused on conditions which are* a barrier to healthy living—such as the nutritional disorders, adverse influences in home and working environment, and the anxieties which bear down on us in our ordinary lives. One of the greatest obstructions to health in many lands abroad was malaria, but I mention it only to commemorate the triumphant researches of Laveran, Ross, and Manson at the close of the nineteenth century, and the founding of the London School of Tropical Medicine in 1899. This is Manson's School. In 1901 the widely distributed disease, beri-beri, so to speak, hung in the balance. Osler describes it as an endemic and epidemic disease of unknown aetiology, but he is familiar with the work of Takaki in Japan, although its success was associated with a wrong theory of cause. He was also aware of the more important findings of Eijkman, and of the Dutch physicians in Java, leading in 1901

comprises these minimal qualitative factors that I am considering.

Publication of the full results of the researches undertaken by Hopkins was delayed until 1912, on account of prolonged and unsuccessful efforts to isolate the actual accessory substances. He did not claim to be the discoverer of 'vitamins' although he might well have done so, but it was certainly the 1912 paper which first awakened an understanding of the physiological significance of these substances The battle was by no means won at the first onslaught, and as late as 1920 there was considerable opposition to the 'vitamin theory' and scepticism about its practical value, especially on the part of a section of the medical profession During the first world war, however, Hopkins was able to exert influence for good, as chairman of official committees on wartime nutrition and famine relief. He often emphasized the inadequacy of a mere calorie figure as an index of nutritional standard. I ought, of course, to add that I am confining myself to the great contribution which Hopkins made to the practice of health. In the realm of pure science his isolation of tryptophane and glutathione were outstanding discoveries.

Shortly after the end of the first world war the work on nutrition was enriched beyond measure by the powerful, logical mind of Edward Mellanby. The application of his studies on rickets to the practice of health illuminated the whole range of child health, and here one should acknowledge with gratitude the contribution of May Mellanby to the prevention of dental caries. For the purpose of this essay I wish to divide Mellanby's investigation into two parts: first, his actual research and its practical results; and secondly, his method of approach, which is of vital interest to the student In the first place it is to be observed that Mellanby's extensive dietary studies on puppies covered a far larger field than the title of the 1921 Report, *Experimental Rickets*, would suggest Mellanby realized from the start that an investigation of rickets could not be confined into a

lished by Gowland Hopkins in 1912 that the significance of dietary prevention, rather than medicinal treatment, was fully grasped. Indeed, the lesson was not learned and applied in the first world war, to our irreparable damage. From 1919 onwards, however, the researches of the Mellanbys demonstrated the essential importance to health of a fat-soluble vitamin in the diet; and cod-liver oil, a rich source of this substance, was master of the situation.

Gowland Hopkins was the father of British biochemistry. He made an unorthodox entry to his profession by apprenticeships, the first of them in an insurance office; but it was through these, especially as assistant to Sir Thomas Stevenson, the Home Office analyst, that he gained a valuable basic knowledge of chemistry. Before the end of the nineteenth century he had invented techniques for the crystallization of pure proteins, and so entered a kind of forbidden ground of investigation. In 1901 protoplasm was all but sacred—containing the inner secret of life. A step forward from this brought Hopkins into the research on accessory food substances which has made his name world-famous. In 1905, with the object of looking for any hitherto unrecognized dietary factors, he fed rats on synthetic foods containing all the known essentials, but in a highly purified form. At first he added meat extracts or yeast to give flavour to the tasteless food, in deference to appetite. He found, however, that rats generally ate the synthetic food quite well without the savoury addition But the extracts powerfully affected nutrition: on the flavoured diet the animals grew well; without the additions they lost weight continuously, although they took the same amount of food. In 1906 Hopkins said in an address to the Society of Public Analysts:

[illegible]

justed to live either upon plant tissues or the tissues of other animals, and these contain countless substances other than the proteins, carbo-

[illegible]

rickets is notoriously a 'healthy' disease; and he points out that catarrhal conditions and great susceptibility to intercurrent infections tend to appear in the wake of rickets. Perhaps we shall never find the secret of the disappearance of another 'healthy' disease—chlorosis—which used to crowd the medical wards when I was a student. The main body of the research, however, is an elaborate experimental study of puppies, and the effect of various diets on the production of rickets. The results stated there have passed into history, and their application has given life and health to a generation of children, both in this country and in others more darkened by the shadow of famine.

Now I should like to turn to method and presentation, for here the younger worker has much to learn from Mellanby. I think that the 1921 Report should be carefully studied as a textbook on this subject. To begin with, the material is set out in a logical sequence, and each section is followed by a short, effective summary. One feels the great value of going step by step with the research worker; it is like being taken into his confidence. For example:

> In addition to the solution of the practical problem of rickets, one other object has been constantly before my mind since it became [illegible]

Next, the methods of examination for signs of rickets are not only described in detail, but objectively presented. A candid summary of the earlier experiments follows: 'It may be of interest to add that the experimental work had continued for about two years before good evidence of differences in the antirachitic action of fats was obtained.' Then comes the main series of experiments, stated simply and with admirable precision, and illustrated by tables, each of which tells its own clear story and no more. Meanwhile, the explanatory text brings the student into the heart of the discussion with the same sense of excitement as if he were present in the same room:

narrow space, like scurvy or beri-beri; and indeed he doubted whether the studies of these conditions had not been too circumscribed. He had grasped completely the need for integration in a study of nutritional defect:

> One point [he says in the 1921 Report] which has come most prominently before my mind as the work has progressed, is the unity of a complete diet and the interdependence of the dietetic elements. This might have been expected. Knowledge has been accumulating in recent years which emphasizes the importance of balanced diets, and we know that to cut out one element of the diet means not only the absence of that element, but also the ineffective action of other elements. For instance, when carbohydrates are removed from the diet, fat is ineffectively oxidized, and there is also good evidence that in their absence animals and plants are incapable of synthesizing proteins from the amino acids. This is a simple instance of the absence of one element upsetting the action of other elements. . . .

After giving a number of examples showing how this interrelationship also holds between these elements and the general metabolism of the body, Mellanby goes on to a masterly summing-up:

> These few instances cannot fail to impress the [illegible] f dietetic diseases that any problem [illegible] may ultimately have to include [illegible] addition to the one that at first [illegible] important It has so happened in this investigation.
>
> While it is evident that a vitamin, probably fat-soluble A, occupies a position of prime importance in the aetiology of rickets, it is undoubted that this vitamin works in a close relationship with the other dietetic elements, and, moreover, with the general activity of the body, so that all the other factors both of diet and environment must be analysed in relation to it

At an early stage of his research Mellanby had to clear [illegible] a considerable accumulation of half-[illegible] e cause of rickets [illegible] By the choice [illegible] and the painstaking application of c[illegible] he was able to show, for example, that lack of exercise (as put forward by Findlay) could not be the prime factor in the development of rickets; so also he effectively demolished the theory that rickets was due to an infection. Incidentally, it is delightful to notice his shrewd paradox that

rickets is notoriously a 'healthy' disease; and he points out that catarrhal conditions and great susceptibility to intercurrent infections tend to appear in the wake of rickets. Perhaps we shall never find the secret of the disappearance of another 'healthy' disease—chlorosis—which used to crowd the medical wards when I was a student. The main body of the research, however, is an elaborate experimental study of puppies, and the effect of various diets on the production of rickets. The results stated there have passed into history, and their application has given life and health to a generation of children, both in this country and in others more darkened by the shadow of famine.

Now I should like to turn to method and presentation, for here the younger worker has much to learn from Mellanby. I think that the 1921 Report should be carefully studied as a text-book on this subject. To begin with, the material is set out in a logical sequence, and each section is followed by a short, effective summary. One feels the great value of going step by step with the research worker; it is like being taken into his confidence. For example:

> In addition to the solution of the practical problem of rickets, one other object has been constantly before my mind since it became evident that a vitamin was bound up in the aetiology of this disease. Our knowledge of vitamins has up to the present depended upon work somewhat limited both in its nature and the number of species of animals investigated, and it was hoped that the use of the dog would not only lead to results more directly applicable to the human being but also would afford a wider basis for considering the part played by these substances in general nutrition

Next, the methods of examination for signs of rickets are not only described in detail, but objectively presented. A candid summary of the earlier experiments follows: 'It may be of interest to add that the experimental work had continued for about two years before good evidence of differences in the antirachitic action of fats was obtained.' Then comes the main series of experiments, stated simply and with admirable precision, and illustrated by tables, each of which tells its own clear story and no more. Meanwhile, the explanatory text brings the student into the heart of the discussion with the same sense of excitement as if he were present in the same room:

In the earlier experiments on the action of fats in rickets, I had not recognized that the amount of bread eaten by the puppies was a crucial point. Up to this time all factors in the diet were controlled except the bread, which was varied according to the appetite of the animal. This fact deprives the earlier work of some of its quantitative accuracy ...

And again:

In most of the later series of experiments, meat was put into the diet in order to induce the puppies to eat up completely each day's ration This plan ensures more uniform results in comparable experiments, but has the drawback of reducing the rachitic changes It is a question of choice between producing very bad rickets but less uniformity for comparison, and more uniformity and less rickets.

The next stage leads us to consider the various hypotheses which have been brought forward from time to time to account for the development of the disease. These are examined in turn, quite objectively, but the findings are expressed so humanly as to give the student a quiet feeling of satisfaction as they are pressed home. Here is an example of what I mean:

It is impossible for me to believe that the drastic treatment of J. Koch to produce rickets in puppies, in which work he injected intravenously cultures of streptococci into the animals, can have any bearing on the disease of rickets in children.

On the more strictly physical side men of science made an equally remarkable contribution to healthy living. One of the outstanding figures both in personality and in the vigorous application of the scientific method to the ordinary things of life was Leonard Hill who, as Greenwood says, was the real founder of applied physiology in England Hill's work on ventilation and more particularly his katathermometer as a measure of the way in which the human body reacts to changes in its physical surroundings brought scientific measurement into the strongholds of superstition and unchecked opinion. The effect of his work has been characteristically described by Greenwood in an essay written in 1931 Here is an extract

... That a bit of glass tubing, which merely enables one to measure how long it takes a little coloured fluid to cool from 100° F. to 95° F., should be of much greater value in the practical control of

atmospheric conditions than a host of elaborate and expensive machines, may seem strange to people who mistake expensive elaboration for efficiency. The point is that Hill's katathermometer does give us a measure of the way in which the human body will react to external conditions, information which can be checked against the

lation and the results we experience, and are familiar with the *psychological attitude of the experts—reminiscent of the shoemaker* who practically tells us that his boots are perfect and we ought not to have such awkward feet. Hill has changed all that.

These investigations led to an enormous quickening of interest in applied physiology not only in the ordinary life of the citizen but also in the factory and workshop, and it was in many ways the corner-stone of later investigations into the environmental conditions of the industrial worker. The personality and vigour which lay behind all Hill's work deserves the most careful study by the student and I can heartily recommend the Report to the Medical Research Committee in 1919 on the science of ventilation and open-air treatment, as well as the works which Hill published in later years His writing is always vivid, and it carries with it colour and freshness as though he were infusing life into the very instruments with which he made his observations. Let me take a few simple examples to show you what I mean:

The observations of Pembrey show that he and four soldiers marching in drill order on a moderately warm day (dry bulb 69° F., wet bulb 59° F.) lost more water and retained more water in their clothes than on another similar day when they worked with no jacket

absurd that on a hot summer day boy scouts should march with a coloured scarf knotted round the neck, or invalid soldiers in Egypt wear a button-up shirt and a red tie. Nothing should be worn for ornament or smartness which increases the difficulty of losing heat

In the earlier experiments on the action of fats in rickets, I had not recognized that the amount of bread eaten by the puppies was a crucial point. Up to this time all factors in the diet were controlled except the bread, which was varied according to the appetite of the animal. This fact deprives the earlier work of some of its quantitative accuracy. . . .

And again:

In most of the later series of experiments, meat was put into the diet in order to induce the puppies to eat up completely each day's ration. This plan ensures more uniform results in comparable experiments, but has the drawback of reducing the rachitic changes. It is a question of choice between producing very bad rickets but less uniformity for comparison, and more uniformity and less rickets.

The next stage leads us to consider the various hypotheses which have been brought forward from time to time to account for the development of the disease. These are examined in turn, quite objectively, but the findings are expressed so humanly as to give the student a quiet feeling of satisfaction as they are pressed home. Here is an example of what I mean

It is impossible for me to believe that the drastic treatment of J. Koch to produce rickets in puppies, in which work he injected intravenously cultures of streptococci into the animals, can have any bearing on the disease of rickets in children.

On the more strictly physical side men of science made an equally remarkable contribution to healthy living. One of the outstanding figures both in personality and in the vigorous application of the scientific method to the ordinary things of life was Leonard Hill who, as Greenwood says, was the real founder of applied physiology in England. Hill's work on ventilation and more particularly his katathermometer as a measure of the way in which the human body reacts to changes in its physical surroundings brought scientific measurement into the strongholds of superstition and unchecked opinion. The effect of his work has been characteristically described by Greenwood in an essay written in 1931. Here is an extract:

. . . That a bit of glass tubing, which merely enables one to measure how long it takes a little coloured fluid to cool from 100° F to 95° F., should be of much greater value in the practical control of

physical health and physical efficiency of workers in munition factories and workshops'.

This active group passed through some vicissitudes and in 1921 it became a branch of the Medical Research Council, with broader terms of reference. It is interesting to note the early limitation to *physical* health and efficiency: the National Institute of Industrial Psychology was established in 1921, and since then the psychological aspect of working conditions has received increasing attention, and a store of new knowledge has been applied to the practice of health. To return to the war, the nation as a whole was beginning to see that overwork, monotony, and especially the lack of any sort of rest or change either through holidays at the week-ends or by means of breaks during working hours, were undermining the health and happiness of the great army of people who were contributing through industry to the war effort. Feelings and opinions, however, are not enough. It was for H. M. Vernon, acting as an investigator for the Industrial Fatigue Research Board, to put the study of working conditions on a scientific basis. Using the experimental method he made a careful inquiry with the object of measuring fatigue and its effect on working efficiency. He examined the records of output in relation to daily and weekly hours of work, comparing one industry with another, and he thus showed by experiment the limitations of human capacity as determined by long hours, monotony, and sheer weariness. Vernon then went on to examine possible remedies, studying particularly the effect of the introduction of shorter spells of working with brief rest periods, and of free time on Saturday afternoons and Sundays. He was able to show, for instance, that a good scheme of regular rest periods might be expected to increase output up to 20 per cent without any increase in fatigue; and again, that the effect of abolishing Sunday labour produced, under favourable conditions, a substantial increase in total output in spite of the loss of eight hours' work. Sunday labour was uneconomical not only because of the physical strain involved but also from the monotony of work week after week without any relaxation

The next subject for study—which has been of great importance to the practice of health—was the frequency and cause of

and keeping down the body temperature. The avoidance of fatigue of the heart, the power to march, and the efficiency depends on prevention of heat stagnation.

And again the vivid contrast he makes between the slavery and confinement of the city worker and the perfect fitness of a free life:

The streets of our cities are trod by hosts of men and women ugly of complexion, ill-grown, weaklings, withered, and worn, incapable of strenuous muscular action at a time which should be their prime, developed in cunning, fear and sensitivity to pain by a mean struggle for existence in mean streets. Contrast these with the wild ass of Sind, whose skin shines with the wonderful gloss of perfect condition, and whose strength and fleetness is such that to ride him down a horseman has to take in relay seven fresh horses. 'The wild asses snuffed up the wind like dragons.' Who hath sent out the wild ass free and men into slavery?

And finally let me take the characteristic commentary of a man who has had to fight for science and knows his enemies thoroughly

It is manifest that some people are acutely affected by draughts, and therefore draught becomes popularly considered the chief cause of colds Nursery and family tradition rules habits, and tradition is usually at least half a century behind science. The room is shut up. Science may beat against the door, but tradition holds it fast.

I think one can sum up Leonard Hill's contribution to the science of ventilation by saying that he gave a great impetus to the study of the physiological basis by the invention of methods of measurement and of instruments capable of giving comparable data.

This leads us to one of the most considerable advances in the practice of health, which arose indirectly from munition work during the first world war. At that time even the most hardened industrialist was becoming uncomfortably aware of the risk of losing output as a result of long hours and fatigue. The serious extent of sickness and lowered efficiency among munition workers led to the appointment of a special committee in September 1915, 'to consider and advise on questions of industrial fatigue, hours of labour, and other matters affecting the

its position hereafter among some of its foreign rivals who already in that respect have gained a present advantage.

Since that time we have had the benefit of learning, through the pioneer work in the psychological field of such investigators as Culpin and May Smith, what a great part is taken by mental health in the balance of working conditions. I am not so much concerned, however, with the practical conclusions of these experimental studies as with the application of the scientific method. Many of the matters considered were in fact capable of direct measurement, but little had been done; others required subjective estimations, and in the hands of experienced workers even these acquired scientific precision through controlled experiment.

I have said enough, perhaps, in a short space to indicate the enormous debt we owe to the physiologists of the first quarter of this century. At any rate, we must now turn to another aspect of development—the field of mental health. It is worthy of notice that Osler, in the fourth edition, devotes little space to functional disorders of the nervous system, but one must in fairness remark that the psychoses were outside the scope of his book. He deals in some detail, however, with such matters as epilepsy, hysteria, and the traumatic neuroses. Neurasthenia, which includes

interesting discussion as to whether it is necessary to postulate spasm in the diseased coronary vessels, or whether the symptoms can be fully accounted for by simple ischaemia on effort, cold, or emotional excitement. The overwhelming preponderance of men as victims is carefully noted, with figures of 237 men to 42 women in his series.

Otherwise practically the whole of what we should describe broadly as the psychoneuroses is a book as yet unopened. In this respect it is striking to observe the development of Osler's interest in the functional disorders as one edition of his work passes to another. In the eighth edition, for example (published in 1913), he is already struggling with Freud's methods and seems at least half-convinced of their value. In many ways he has acquired

accidents. Vernon was able to show the malign effects of fatigue on the accident rate, and in a most interesting way the effects of lighting and temperature. From carefully drawn thermograph charts he demonstrated that a general level of between 65° and 69° F. could be regarded as optimum, and that the accident rates rose both above and below these figures. In a similar way the results of bad lighting were measured in terms of output and quality of work.

In its final report, issued in 1918, the Health of Munition Workers Committee urged the need 'to make arrangements without delay for a national scheme for industrial medical research, and to accord fuller recognition to the importance of industrial hygiene'. The research board set up as a result of this recommendation concentrated their attention on four main problems:

1. The fundamental question of shorter hours of labour, from a political and economic (as well as a health) point of view; for this was said to lie near the root of the whole labour problem.

2. The social and economic conditions of women's labour, over and above the issues of health which were dealt with by the Committee, for the health conditions of women are inseparable from their social conditions.

3. The solidarity of industrial society—the interdependence of employer and workman, which is closely related to the whole issue of the status, health, and physical equipment of the worker.

4. The title of the worker to an effective voice in regard to the conditions under which he works.

The Committee concluded its discussion of the relation of fatigue and ill-health to industrial efficiency with this confession of faith:

> The national experience in modern industry is longer than that of any other people. It has shown clearly enough that false ideas of economic gain, blind to physiological law, must lead, as they led through the nineteenth century, to vast national loss and suffering. It is certain that unless industrial life is to be guided in the future (i) by the application of physiological science to the details of its management, and (ii) by a proper and practical regard for the health and well-being of our workpeople in the form both of humanizing industry and improving the environment, the nation cannot hope to maintain

or of the active belief in the assurance of the physician that the precious boon of health is within reach.

It is not for me to assess the relative value of the contribution of the analytic method—the contribution made by Freud and his followers—and that of other workers in the field of mental health, especially after the first world war; this in itself is an important chapter in the history of human endeavour. Let me at least point out the great enlargement in the concept of mental health as a positive ideal to be achieved as opposed to mental disorder, a disease to be treated. This, rather than the claims of rival methods of treatment, constitutes a revolution in the practice of health It is one of the hopes of a sick and frightened world.

Until the year 1930 the law dealing with persons of unsound mind was obsessed with ideas of detention to the virtual exclusion of prevention and treatment. Moreover, a sharp distinction was maintained between the private patient and the 'pauper lunatic' long after these invidious social prejudices had died out in other spheres of medicine. In the first quarter of the present

many types of early disorder the patient was able and anxious to co-operate with the physician in treatment. The Report of the Royal Commission of 1926 begins a new chapter with an important statement of principle:

Legislative changes have now implemented these recommendations to a large extent, the greatest benefit being the ready admission of voluntary patients to mental hospitals. But the new

considerable understanding of the significance of Freud's theories but he is still doubtful and rather detached when it comes to their application to therapeutics. In speaking of the sexual causes of neuroses, for example, he weaves a little fantasy of his own.

> Undoubtedly the part played in the production of hysteria and allied neuroses by sexual factors is of the first importance. As already
>
> basis of the psychoneuroses. Repressed as they have to be in so many in our modern civilization, without normal outlet, the thought formations, retained in the unconscious state, express themselves by means of somatic phenomena—the objective features of hysteria and neurasthenia. *Cherchez la femme* is a safe rule in investigating a neurotic case. Freud may have ridden his hobby too hard, particularly in the insistence upon the importance of infantile sexuality, but in recognizing the role of the younger Aphrodite in the lives of men and women he has but followed the great master, Plato, who saw, while he deplored, the havoc wrought by her universal dominance.

When he comes on to psychotherapy the division in his own mind is more clearly marked; and as usual, he covers the cracks by painting his own charming brightly coloured words:

> The use of religious ideas and practices may be most helpful, and this has come into vogue in various forms, as Christian Science, Emmanuelism, Mental Healing, etc. It is an old story. In all ages,
> has
> an
> he
>
> that the personal character of
>
> or Ste Anne de Beaupré. Three elements are neces-
> faith—
> saints,
> well, a
> physician Secondly, certain accessories—a shrine, a sanctuary, the

XV

Research on Ageing

E. C. DODDS

WRITERS and speakers on the subject of old age for the last 2,000 years have turned to Cicero's *De Senectute.* Whilst this may have been a wise move up to 500 years ago there is no doubt that the modern speaker or writer can gain little or no help with his problem today from this source. It is interesting to commence with the first paragraph of the *De Senectute* addressed as it is to his friend who is retiring on account of old age, probably at that time an age of between forty and fifty:

> However, at the present, I have determined to write something on old age to be dedicated to you, for I fain would lighten both for you and for me our common burden of old age, which, if not already pressing hard upon us, is surely coming on apace, and yet I have certain knowledge that you, at all events, are bearing and will continue to bear that burden, as you do all others, with a calm and philosophic mind. But when I resolved to write something on this theme you continually came before my mind as worthy of a gift which both of us might enjoy together. To me, at any rate, the composition of this book has been so delightful that it has not only wiped away all [illegible]

This work was written in 54 B.C. and the expectation of life in Rome at that time was certainly not 30 years at birth. Forty to fifty was considered quite an exceptional survival, whilst anyone over 60 approached immortality. It is little to be wondered at that in earlier times old age as judged by these standards was

spirit has been carried much further than this into the realm of prevention, and the promotion of mental health has been fostered through skilful teaching in schools, in industry, and in clinics for guidance and after-care. It has been suggested in some quarters that there is a risk of overstressing the mental aspect of sickness—an epidemic of cults and creeds. This danger is ever present in the art of medicine, but it cannot be overcome except by providing more and better services; and we know that the services for the care of mental patients of every kind are tragically insufficient. Our plain duty is therefore to encourage more and more research and experiment in the field of prevention and in the promotion of health. This cannot be achieved simply by the provision of more beds or even by the multiplication of specialist clinics. The real contribution to the practice of health can be made effective only through the media with which everyone is familiar: the general medical practitioner, the health visitors, the press, and the radio: in other words, the communications that reach the homes of the people.

REFERENCES

Final Report of the Health of Munitions Workers Committee, April 1918. H.M.S.O.
GREENWOOD, MAJOR (1931). The Work of the London School of Hygiene and Tropical Medicine, *Jnl of the Royal Society of Arts*, May 1, p. 542.
HILL, L. (1919). *The Science of Ventilation and Open Air Treatment*, Pt. 1. Medical Research Committee Report H.M S O.
MELLANBY, E (1921). *Experimental Rickets*. Medical Research Council. H.M.S O.
NEEDHAM, J. (1949) *Hopkins and Biochemistry* Heffer, Cambridge.
OSLER, W. (1901). *Principles and Practice of Medicine* (4th edn.). Appleton, New York and London.
Report of the Royal Commission on Lunacy and Mental Disorder (1926). H.M.S O
VERNON, H. M. (1940) *The Health and Efficiency of Munition Workers*. Oxford University Press, London

Medicina Gerocomica:

OR, THE

Galenic ART

Of PRESERVING

Old Mens Healths,

EXPLAIN'D:

In Twenty CHAPTERS.

With an APPENDIX, concerning the Uſe of Oils and Unction, in ſome Diſeaſes. And a Method, from a *Florentine Phyſician*, of curing *Convulſions* and *Epilepſies*, by external Operation.

By Sir JOHN FLOYER, Kt. of *Lichfield*, M.D.

The Second Edition, corrected. *To which is added,* A LETTER to the Hon^ble Mr. *Ch*—— *St*——: Concerning the Regimen of the Health of the Younger Years and Adults, as *Galen* has deſcrib'd them.

Pugnandum tanquam contra morbum, ſic contra Senectutem, Cicero de Senectute.

Calida lavatio & Pueris & Senibus apta eſt vinum dilutius Pueris, Senibus meracius. Celſ. de re Medic. lib. 2.

LONDON:

Printed for J. ISTED, at the *Golden-Ball*, between St. *Dunſtan's* Church and *Chancery-lane* End, in *Fleetſtreet*. MDCCXXV. Price ſtitcht 2 s. 6 *d*. bound 3 s.

FIG. 1 Title page of the second edition of Floyer's *Medicina Gerocomica*.

greatly respected and that tribes and society of earlier nations were usually ruled by 'elders'. In Sparta there was the conclave of elders known as the gerousia, the word being derived from the Greek word γέρων, meaning old man. In order to be elected to the gerousia a candidate had to be over 60 and if we are to believe the calculated survival rates there can have been very few applicants for the posts.

Coming to more modern times a study of the organization of the tribes of aboriginals in Australia again shows this rule by the elders. Here again the word meant someone over 40 and it has been repeatedly stated that an aboriginal of over 50 living in his original surroundings is certainly an unusual occurrence. It is not proposed here to enter into anything like a detailed historical review; one only wishes to mention one or two classical works. Up to the Middle Ages most works on the problem of ageing were of a philosophical character. The problem as one of health and physical well-being appeared to escape the early authors, who were deeply influenced by Cicero and the classical writers. In the thirteenth century we find the first book on the problem of old age considered more from the physical point of view. This was written by Roger Bacon (*The Care of Old Age and Preservation of Youth*, translated by Richard Browne, 1683). Whilst it is always unwise, particularly at an open lecture such as this, to make a definite claim for priority, I feel that one can put forward the name of Sir John Floyer as the writer, if not of the first English book, certainly of one of the first dealing with the study of old age from a health point of view. Fig. 1 shows the introduction of the title page of the second edition of Floyer's *Medicina Gerocomica* published in the year 1725. Floyer was knighted for his contribution to medicine and he had an extensive practice in Lichfield.

The first volume of this work, which is in the library of the Royal College of Physicians, forms most interesting reading and has a remarkably modern flavour in many of its passages. The author's tenets are really the taking of care in exercise and diet and so forth and the object is to put as little strain as possible on the mechanism of the ageing body. He also recommends treatment with baths and suggests alternating hot and cold rubs in

Medicina Gerocomica:

OR, THE

Galenic ART

OF PRESERVING

Old Mens Healths,

EXPLAIN'D:

In Twenty CHAPTERS.

With an APPENDIX, concerning the Use of Oils and Unction, in some Diseases. And a Method, from a *Florentine Physician*, of curing *Convulsions* and *Epilepsies*, by external Operation.

By Sir JOHN FLOYER, Kt. of *Lichfield*, M.D.

The Second Edition, corrected. *To which is added,* A LETTER to the Hon[ble] Mr. *Ch---- St----*: Concerning the Regimen of the Health of the Younger Years and Adults, as *Galen* has describ'd them.

Pugnandum tanquam contra morbum, sic contra Senectutem, Cicero de Senectute.

Calida lavatio & Pueris & Senibus apta est vinum dilutius Pueris, Senibus meracius. Cels. de re Medic. lib. 2.

LONDON:

Printed for J. ISTED, at the *Golden-Ball*, between St. *Dunstan's* Church and *Chancery-lane* End, in *Fleetstreet.* MDCCXXV. Price stitcht 2 s. 6 d. bound 3 s.

FIG. 1. Title page of the second edition of Floyer's *Medicina Gerocomica*.

order to tone up the patient. He points out the dangers of constipation and suggests the use of shale and mineral oils towards this end.

From then on a large number of treatises appeared, but it is not our object here to attempt a literary review. By the end of the nineteenth century the problem of the study of old age had become clearly recognized as an intellectual task imposed on the *existing generation* and the science of gerontology was born. This was meant to indicate the whole field, but particularly with reference to a study of the nature of the ageing process. Later the term 'geriatrics' was introduced in order to distinguish the more practical side of the problem. Geriatric medicine was intended to convey therapeutic measures in use in the treatment of old age as a condition in itself and illnesses associated with old age.

Only the briefest reference will be made here to the sociological problems involved in this study and only enough will be said in order to concentrate attention upon the urgency of the problem. As we have said at the beginning of this lecture, at the time that Cicero wrote his classical treatise on old age, the expectation of life at birth was probably under 30—round about 28—whilst very few people survived beyond the age of 40. This state of affairs continued past the Middle Ages in this country, and it was only with the introduction of the new hygiene procedures that accompanied the industrial age in the nineteenth century that an increase in the expectation of life was shown. By the year 1900 there had been a substantial increase in the expectation of life at birth and from 1900 to 1949 this steadily increased.

It is easy to see that an increase in the expectation of life must alter completely the composition of the society with regard to ... shown in the ... l diagram- ... to over 80 ... line. The figure on the extreme left shows that the diagram is triangular, that the base is at the 0–4 age group and that it remains fairly constant up to 20–24, but after 30 the sides begin to shelve in so that at 60 the number of people alive has fallen very con-

siderably. If we turn to the 1940 diagram, the middle one, we can see the effect of increasing the age of survival on the distribution of the population at any given moment. To start with, the base at 0–4 is nothing like so large and this is due of course to the fact that the numbers are diluted with those of a higher age group. It bulges out reaching a maximum at 15–19 and then slowly closes in to the 65–69 age group. If we turn to the

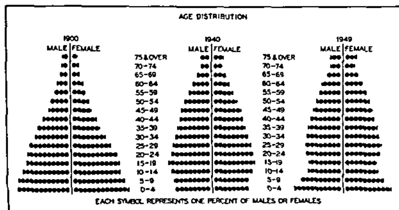


Fig. 2. Age distribution trends of the population of the United States

diagram on the extreme right we see that in 1949, that is to say 50 years after the first one, the whole picture has become completely changed, the diagram has become much narrower and the sides are tending to be parallel to the centre line particularly round about the period from 30 to 60. This means that the younger people are being diluted with the survivors who owe their existence to the increase in life expectation due to therapeutic and other measures. This can be put in another way and is shown very clearly in the next diagram (Fig. 3). Here one can see that in 1879 the expectation of life at birth was 34 whereas by 1950 it had risen to about 68. Dublin (1941) and his co-workers, by studying all available data, have drawn a very inter-

esting curve of life expectancy at birth from the earliest time to the present day and this is shown in Fig. 4. Here we see that in the year 500 B.C. the expectation at birth was under 20, and at the present day is approaching 70. Admittedly though the data on which the early part of this curve is based are somewhat inferential to say the least of it, there can be little doubt that it is substantially correct.

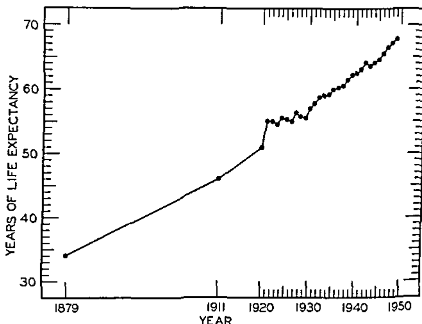
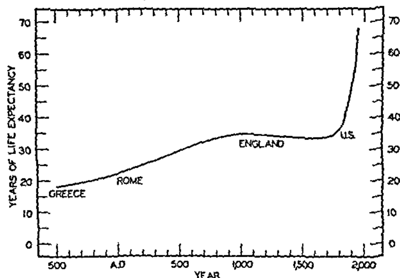


FIG. 3. Life expectancy, 1879–1950 (average age of deaths). Data from *Statist. Bull Metrop. Life Insce. Co.* 27, 1–7, January 1950. (Repr. from Shock, 1951, *Trends in Gerontology*)

This increase in the survival rate is due to a number of factors, of which undoubtedly the greatest is the introduction of therapeutic measures. Fig. 5 shows very clearly the effect of some of the introductions of new remedies on the death rate per 100,000. Here we can see that the introduction of serum treatment for pneumonia in the late twenties caused a steep decline in the death rate whilst the introduction of the so-called sulpha drugs in the middle thirties produced another steep decline and the advent of penicillin pushed the figure down considerably more.

Another interesting diagram (Fig. 6) shows again the relationship between survival and disease. The deep black shows the male and female population of the United States in a survey in 1940–1 and shows their life expectancy as determined from the survey. The single shaded area on top of these columns shows what the extra expectancy would have been if one could have eliminated cardiovascular and renal disease, and one can



Gerontology.)

see that it is something like ten years through the whole series. If one could eliminate cancer, as shown by the double shading, then another series of years would be added. The import of this surely indicates that at the age of 70, if one could exclude cardiovascular disease and cancer the added expectancy of life would be 20 years—truly a terrifying thought to those concerned with the care of such a population. One final point in the general consideration· up to the present we have only studied gross or global survivals indicating that as time goes on and more therapeutic means become available the population at any given year

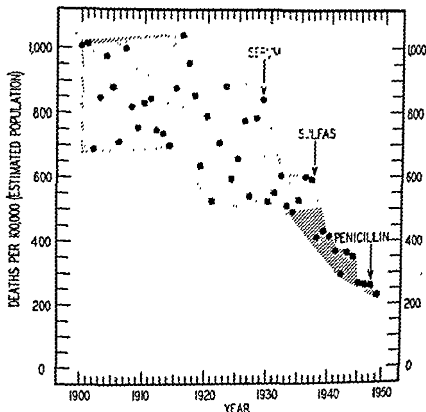


Fig. 5. Death rates for pneumonia and influenza for age group 65 years and over; death registration states, 1900–48 (ratio per 100,000 estimated mid-year population). Arrows indicate the introduction of new therapeutic agents Data from National Office of Vital Statistics, U.S. Public Health Service.

will alter in such a way that there will be an increase in old people.

Yet another aspect for the medical planner is the effect that this change in population age has on the type of disease. This is clearly shown diagrammatically in Fig. 7. The diagram attempts to show the changing nature of the diseases responsible for death over the period 1900–40. If we start at 1900 with the five great causes of death, namely heart disease, intracranial lesions of vascular origin, pneumonia and influenza, nephritis and finally cancer, we see a fairly close packing for the first three with a

considerable drop off in the case of nephritis and cancer at the last. This is summarized in the circle on the right which shows that roughly half the deaths are caused by these five diseases. If we go up the column to the year 1910 we find that heart disease

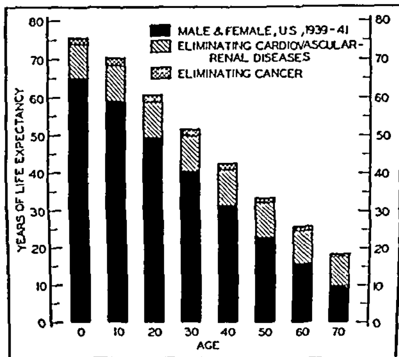


FIG. 6 [illegible]

has leapt ahead as a cause of death whereas intracranial vascular lesions are roughly the same; influenza is last and cancer is on the increase. This *result must be read together with the* population studies to which we have just referred and there can be no [illegible]

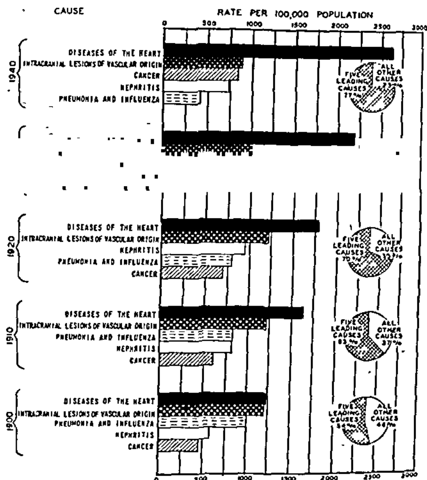


FIG. 7. Rank order of five leading causes of death for adults 65 years and over. U.S. death registration states, decennial years 1900–40. Data from National Office of Vital Statistics, Special Reports, 16, 339 (1944). (Repr. from Thewlis, 1946, *The Care of the Aged*)

more and more of the circle until in 1940 over 75 per cent of the deaths are due to the five main causes. Here we see a steady increase in diseases of the heart which by the top has doubled itself as compared to 1900. We notice also that cancer has now increased to practically double the 1900 rate and has become equal almost to that of the intracranial lesions. Pneumonia and influenza have now become the smallest.

This study shows very clearly that the change in the distribution of age groups in population will bring with it an entirely new incidence of disease and that the medicine of the future will obviously be concerned with the chronic diseases of cardiovascular origin and particularly with cancer until such a time as these can be eliminated.

The picture is, therefore, an alarming one and requires detailed consideration by every branch of society. What again is paradoxical is that the more advances we make the more difficult will become the problem. For example, we have already pointed out that if a cure for cancer could be evolved then even with our present knowledge of cardiovascular disease the expectancy of life at the age of 70 would be 20 years and it is certainly not, we trust, outside the range of possibility that the cancer problem might disappear with the same rapidity as that of pneumonia, and those previously fatal conditions that now respond to penicillin and other antibiotics

So much for the introduction of this lecture, the object of which is to attempt to focus the attention on the urgent necessity for every type of consideration for the problem of old age. We have shown by this brief excursion into the statistics that whether we like it or not we are going to be faced with an increasing number of old people. The main object of this present discussion is to consider the ageing process itself as apart from diseases in old age which is not my subject. What is the nature of the ageing process? If a person can be kept free of disease he or she nevertheless goes through this period of senescence with its characteristic clinical features, so well known to the medical profession and the laity alike. That this process of senescence advances more rapidly in some than in others is a matter of common experience, but that it will appear in all humans provided they live long enough is a reasonable statement. A great deal has been written on theories concerning the ageing process, but most is speculation based on inadequate experimental work and I do not propose even to review the various theories advanced from time to time to account for the senescence process. We can look through the other realms of nature to see whether senescence occurs elsewhere than the human race. A very full review

of this will be found in Professor Cowdry's excellent work (1942) and as he points out we can get very little help from comparative studies. For instance, the realm of botany really does not help us at all in view of the entirely different types of life in plants, although it is well known that certain specimens such as trees will continue to grow for very long periods and then show signs of atrophy and will die apparently spontaneously without being attacked by fungi or other lethal factors in the life of trees. Again the study of insects does not help as their lives and reproduction are so entirely different from mammals that even with the broadest mind no useful inspiration can be found. Unicellular animals are, so to speak, immortal as they reproduce by fission and are not, therefore, in any way comparable.

Studies of the vertebrates show that in those instances where violent death is avoided senescence with its systemic classical features intervenes in a manner very similar to that occurring in the human subject. Studies of the experimental animal have revealed that if one removes tissues from vertebrates it is possible to cultivate them by the methods of tissue culture almost indefinitely and, therefore, the failure associated with senescence is some general systemic action and may not be concerned with specific failure in the tissues.

I think that the only discussion that can be of any value to us at the present time with our very limited knowledge is an attempt to find out whether senescence is due to some systemic process of failure or whether it is due to disintegration of individual structures.

The hypothesis that the ageing process is due to some general systemic or controlling factor leads one to consider the endocrine system as the main possibility. Ever since the appreciation of the rôle of the internally secreting organs, they have been associated in both lay and professional minds with rejuvenation. That the pituitary may be associated with the ageing process has long been a favourite thesis and a study of Simmonds' disease would tend to support such an idea. This condition, described by a German clinician in the early part of this century, is characterized by a sudden ageing of the person. The face becomes lined, the hair falls out and a typical picture of premature ageing

is presented. Further investigation shows emaciation, loss of hair and suppression of the menstrual periods. When the condition occurs in men atrophy of the testes and external genitalia occurs with of course complete loss of potency. Simmonds was able to connect these clinical findings with a degenerative lesion of the anterior lobe of the pituitary gland. It is also interesting to note that with improvement in the technique of cerebral surgery, post-operative Simmonds' disease is encountered quite frequently in patients who have had tumours of the pituitary removed. Again partial atrophy of the anterior lobe is said to occur following certain cases of toxaemia of pregnancy and this is attended by the partial appearance of the symptoms of *Simmonds' disease.*

If we turn to the experimental side, we now know that hypophysectomy does not produce in the experimental animal the typical signs of premature ageing. A competent worker can now hypophysectomize the rat and if the operation has been successful a large survival rate can be obtained. The animals do not lose weight, but of course they show a suppression of the secondary sex characteristics in both sexes due to the fact that the gonadotropic hormones are no longer available from the anterior lobe of the pituitary to keep the testes or the ovaries in their normal functioning condition.

Again, castration of the male or female animal, whilst producing the characteristic effects associated with either non-development or suppression of the secondary sexual characteristics, does not induce the senescent process. This is well known in the case of human beings. In the case of women with bilateral cystic disease, if bilateral ovariectomy is performed before the onset of puberty a state of eunochoidism is produced whereas if performed after puberty then a condition similar to the menopause intervenes, but premature ageing certainly does not occur. Again, the castration of the male has been practised in the East for the production of eunuchs and in Europe for producing castrato singers who again appear to age in the normal manner and at the normal rate. Contrary to these single observations it can be stated that arrest of general senescence cannot be obtained by the administration of either oestrogens to the female

of this will be found in Professor Cowdry's excellent work (1942) and as he points out we can get very little help from comparative studies. For instance, the realm of botany really does not help us at all in view of the entirely different types of life in plants, although it [illegible] known that certain specimens such as trees will continu[illegible] [illegible]ds and then show signs of atrophy an[illegible] being attacked by fungi or other le[illegible] the study of insects does not help as their lives an[illegible] are so entirely different from mammals that even with the broadest mind no useful inspiration can be found. Unicellular animals are, so to speak, immortal as they reproduce by fission and are not, therefore, in any way comparable.

Studies of the vertebrates show that in those instances where violent death is avoided senescence with its systemic classical features intervenes in a manner very similar to that occurring in the human subject. Studies of the experimental animal have revealed that if one removes tissues from vertebrates it is possible to cultivate them by the methods of tissue culture almost indefinitely and, therefore, the failure associated with senescence is some general systemic action and may not be concerned with specific failure in the tissues.

I think that the only discussion that can be of any value to us at the present time with our very limited knowledge is an attempt to find out whether senescence is due to some systemic process of failure or whether it is due to disintegration of individual structures.

The hypothesis that the ageing process is due to some general systemic or controlling factor leads one to consider the endocrine system as the main possibility. Ever since the appreciation of the rôle of the internally secreting organs, they have been associated in both lay and professional minds with rejuvenation. That the pituitary may be associated with the ageing process has long been a favourite thesis and a study of Simmonds' disease would tend to support such an idea. This condition, described by a German clinician in the early part of this century, is characterized by a sudden ageing of the person. The face becomes lined, the hair falls out and a typical picture of premature ageing

ing age. Aub and Du Bois (1917) tried but were unsuccessful in obtaining healthy volunteers between 90 and 100 years old. After much labour they made a study of six men between 77 and 83 years. They found the average heat production, as measured in the heat calorimeter, was 35 1 calories per square metre per hour as compared with 39 5 calories for men between the ages of 20 and 50, or a decrease of 12 per cent. The findings of Aub and Du Bois have been criticized by Benedict and Root (1934) on the grounds that all the men were senile, but it was not made clear in their criticism whether senility raised or depressed the heat output. It is difficult to believe it could do either. Benedict and Meyer (1932) made a study of the basal metabolic rate of twenty-three women between the ages of 66 and 86 and they again confirmed that the heat output decreases with advancing age. They made the interesting observation that those subjects who were the most active had, contrary to popular belief, the lowest metabolisms.

Several interesting single observations have been carried out on nonagenarians—all men. Benedict (1928) in America measured one subject, while in England, Wolf of Addenbrooke's Hospital, Cambridge, measured Sir James Crichton-Browne. The latter must have been an amazing man for his years, for when in his ninety-fourth year he was active and still writing letters to *The Times*—one on 'The Canadian Boat Song' on 12 February 1934, and another three days later on 'Slaughtering of Cattle'. Sir James died on 21 January 1938, in his ninety-eighth year.

Very interesting observations have been reported on the basal metabolic rate with advancing age by Magnus-Levy (1942). At the age of 26 his own heat output per square metre per hour, when measured in Zuntz's laboratory, was 38·1 calories with a basal metabolic rate of −4 (Aub and Du Bois' standards), while at the age of 76 his heat output had fallen to 31 5 calories per square metre per hour with a basal metabolic rate of −11 (Aub and Du Bois' standards).

In summary it may be said that with advancing years the body's heat output declines but there is a great deal of variation. The relatively small number of cases recorded makes the normal

or androgens to the male. Extensive experiments have shown that it is impossible either to arrest or to postpone senescence by their administration. We must admit that whilst the orderly nature of the changes occurring in the ageing process would appear to suggest some co-ordinated centralized mechanism its nature is as yet unknown.

Turning to the second possibility, namely the failure of individual organs and tissues, we are able here to call on extensive researches. First of all let us consider the question of general metabolism. The basal metabolic rate has been the subject of research for very nearly a hundred years. The heat output is best measured for comparative purposes in relation to the area of the body and the basal metabolic rate is expressed as a percentage of the heat output of a normal person. American investigators have made extensive observations on normal people from which the standards are obtained. It is known that age affects the basal metabolic rate. Much more study will be required on the B.M.R. in old age before satisfactory conclusions can be drawn. It is no easy matter getting material for this study, for, by the time the end of the statutory life period of three score years and ten has been reached, diseases, many of them affecting metabolism such as hypertension, heart disease, renal disease, and conditions causing wasting, not uncommonly are attacking or have attacked man. While it is easy to get volunteers for a study of the metabolism of youth, one requires in old age to rely for one's volunteers mainly on so-called 'hospital normal', that is people confined to hospital or a convalescent home. These subjects cannot be considered as representative of the true picture found in healthy old age. Despite these difficulties several reports have appeared on single cases of healthy old age, and there are a few reports where a series of old people have been studied.

Andral and Gavarret (1843) were probably the first to notice that with increasing age there was a decrease in the carbon dioxide output. The real credit for our knowledge of what happens to the basal metabolic rate with increasing age goes to Magnus-Levy and Falk (1899). They studied a group of elderly men and women up to the age of 86 years, and found the oxygen consumption fell (per kg of body weight) with increas-

vestigated the urea clearance of a number of subjects from 40 to 89 years of age and showed that there was a tendency for the urea clearance to decline down to about 55 per cent with advancing age. Similar results were obtained by Davies and Shock (1950) for inulin and diodrast clearances. These results are interesting to contrast with those obtained by McCance (1948, 1950) who studied renal function at the other end of the scale, namely the new-born infant. He showed that the new-born kidney was a relatively inefficient organ, was unable to concentrate and had an efficiency of about one-fifth of the adult kidney. So here it would seem that the ageing kidney is the more useful organ than the neonatal kidney.

Skeletal changes. The increase in the brittleness of the bones associated with advancing age has been known from time immemorial and the intracapsular fracture of the neck of the femur is a classical method of terminating one's existence over a certain age. Extensive investigations have been performed on this osteoporosis of old age. Senile osteoporosis or osteomalacia of the spine has been described by a number of workers including Meulengracht (1938), Black, Ghormley, and Camp (1941), Albright, Smith, and Richardson (1941), Burrows and Graham (1945). These workers have shown that the condition is commoner in women than in men and occurs particularly in the age group of 60–65. A very interesting research was conducted by Ian A. Anderson (1949). This worker, on theoretical grounds, associated the osteoporosis in women with the menopausal syndrome and decided to investigate the effect of administration of an oestrogen with the object of seeing whether it could influence in a positive manner the calcium balance. He used the synthetic oestrogen, dienoestrol, which our laboratory, together with Sir Robert Robinson, produced some twelve years ago. He was able to show that the administration of dienoestrol produced an immediate subjective improvement together with absolutely definite radiological evidence of increased deposition of calcium in the skeleton These changes could not be obtained by the use of calcium and vitamin D alone. It would seem that here is a very practical point and that possibly by the administration of calcium plus synthetic oestrogens to women over the age of 60

standards difficult to compute. The position is well summarized by Du Bois (1936): 'If the general slope of the standards are continued into old age, the Harris-Benedict level is somewhat too low and the Aub and Du Bois and Boothby a little too high.'

If we turn to individual organs the question of the digestive system immediately arises. Gastroentestinal disturbances are common in old people and anorexia with the characteristic glazed tongue is not infrequently seen. A very thorough investigation into gastric function in 100 normal elderly people was made by Daniel T. Davies and T. D. Illtyd James in 1930. This investigation, by far the most complete in this field to that date, was conducted on persons over the age of 60, the eldest being 95 years of age. Great care was taken to exclude persons suffering from intercurrent or associated diseases and we can take this series as representing the picture in 100 normal people of advancing age. This series can be compared very aptly with that of Bennett and Ryle who performed gastric analyses on 100 healthy students in the year 1921. They showed that out of this series only four had achlorhydria. In the Davies and James series 32 of the 100 showed no free hydrochloric acid by the test meal and a re-investigation of these 32 by the administration of histamine reduced the number of cases showing complete achlorhydria to 15 per cent. This shows that achlorhydria tends to increase as age advances but it is surprising that brisk secretion of both acid and pepsin is encountered in so many old people. Davies and James also noted that the atrophic changes of the mucous membrane of the tongue were found with more frequency in those cases showing a diminished secretion of gastric [illegible] normal secretion [illegible] use of [illegible] indeed [illegible] ure of other parts of the tract is a contributory cause

Kidney. Fortunately there exists a very clear-cut method of investigating renal function by means of the various types of clearance test. A search of the literature shows a number of publications describing the application of these tests to study renal function in relation to age Lewis and Alving (1938) in-

compare the 'picture' or 'pattern' at various age levels. We have made a preliminary essay in this direction recently, by comparing the pattern obtained from urine of a group of twelve premenopausal women (ages ranging from 16 to 38 years) with those from a similar group of postmenopausal women (ages 50

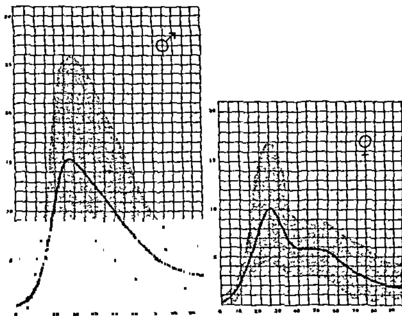


FIG. 8. Urinary output of 17-ketosteroids (as mg/24 hrs) at various ages in males and females. (Repr. from Hamburger, 1948, *Acta endocr. Copenhagen*, 1, 19)

and over) The comparison has been made in each of the fractions (1) ketonic alcohols, (2) non-ketonic alcohols, and (3) ketonic non-alcohols (Fig. 9). We [illegible]

fractions.

In contrast, the output of 'corticoids' shows no dramatic fall with increasing age (Fig 10). Such corticoid estimations must also be regarded as the sum of many components, some associated with mineral-corticoid functions and others with gluco-corti-

the risk of intracapsular fracture of the femur could be eliminated.

The application to men is certainly of interest. It can be shown that the administration of synthetic oestrogens will cause the deposition of calcium in the male skeleton and it would be interesting to know whether similar results could be obtained with testosterone.

In conclusion one would like to make an appeal for intensification of research on the ageing problem. The brief outline that I have given has already shown the number of lines which could be profitably pursued, and in one particular case, namely the osteoporotic changes, actual therapy of a preventive character is indicated. I would conclude with two additional indications. During recent years elaborate techniques have been evolved for the investigation of sterol metabolism by a study of the distribution of urinary steroids. Mainly through the work of Dobriner and his colleagues in America (1942–8) it has been possible to estimate the urinary steroids in a number of normal cases and to study the various constituent members of the group. It is claimed by Dobriner that there is a characteristic urinary steroidal picture for a number of diseases and it is indeed a pity that extensive investigations of the steroid pattern in old age have not been conducted. A certain amount has been done and the following is a very brief summary.

Fig. 8 shows in graphical form the results of the urinary output of 17-ketosteroids carried out on 137 normal male subjects (3–102 years of age) and 127 normal females (2–92 years of age) (Hamburger, 1948). This graph illustrates clearly the falling ketosteroid output with increasing age and serves to emphasize points noted by Heller and Shipley (1951) that (a) when compared with normal young subjects of the same sex the percentage decline in males is more severe than in females, and (b) that in the ninth decade the values are not much lower than in the seventh decade.

These observations are based upon total 17-ketosteroids and this figure conceals the fact that many individual and recognized ketosteroids contribute to this total. Fortunately, methods are now becoming available to resolve these components and to

compare the 'picture' or 'pattern' at various age levels. We have made a preliminary essay in this direction recently, by comparing the pattern obtained from urine of a group of twelve premenopausal women (ages ranging from 16 to 38 years) with those from a similar group of postmenopausal women (ages 50

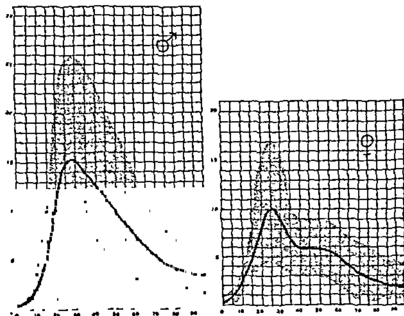


Fig. 8 Urinary output of 17-ketosteroids (as mg/24 hrs.) at various ages in males and females. (Repr. from Hamburger, 1948, *Acta endocr. Copenhagen*, 1, 19)

and over). The comparison has been made in each of the fractions (1) ketonic alcohols, (2) non-ketonic alcohols, and

fractions.

In contrast, the output of 'corticoids' shows no dramatic fall with increasing age (Fig. 10). Such corticoid estimations must also be regarded as the sum of many components, some associated with mineral-corticoid functions and others with gluco-corti-

coids. These findings contrast oddly with the gonadotropin excretion in old age. Heller and Shipley found that the gonadotropin excretion exceeded the normal range in 70 per cent of the aged women studied and in 21 per cent of the aged men and

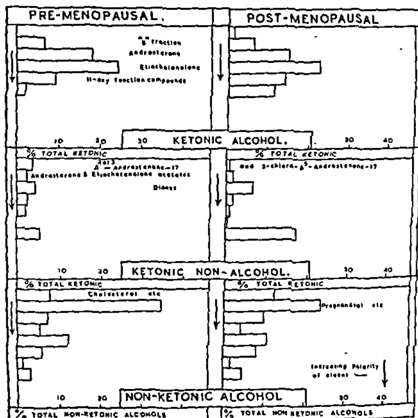


FIG. 9. Urinary steroidal excretion patterns, shown by chromatographic studies, in normal women.

could find no correlation between these levels and the climacteric symptoms. In women, as the postmenopausal interval increased there was a tendency for the gonadotropic titres to be lower than in the early postmenopausal period. Many subnormal values were encountered in the aged men which reduced the overall increase. Despite this evidence of reduced pituitary activity, there is undoubtedly hyperactivity of the

pituitary gonadotropic hormones following the climacteric, at a time when the secretion of the target organs has already begun to fall.

Finally, the use of isotopes opens up an entirely new field for the study of the ageing process. Time does not permit me to go into the dramatic change of viewpoint caused by the introduction of this technique. The work of Schoenheimer (1942) with deuterium and ^{15}N labelled amino-acids has shown conclusively that the proteins of the body are in a perpetual state of dynamic

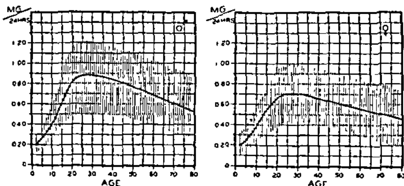


FIG. 10. Urinary output of corticoids (as mg/24 hrs.) at various ages in males and females.

change, and that the protein structures of the body are constantly being broken down and built up again Thus he was able to show that the liver protein of the rat is rebuilt in seven days. It has been calculated, mainly from the investigations of Rittenberg, that in man the whole of the proteins of the body are turned over in 80 days So far as I know no work has been done on the effect of the ageing process on this dynamic equilibrium of the proteins of the body and surely here is a field simply asking to be investigated.

It may well be that by the application of these methods some new facts may emerge with regard to the ageing process. Unfortunately the funds available for research in this country on the ageing process are very limited. The Society for Research on Ageing has been generously supported by Lord Nuffield and it is to be hoped that this and other organizations will be able to

collect funds so that these researches can continue. As we showed in the beginning of this lecture the number of old people will increase steadily and it behoves us to make a maximum effort to see that the health of this group of the population is maintained at the highest level.

As we began with Cicero, so might we end:

> Again, if we are not going to be immortal, nevertheless, it is desirable for a man to be blotted out at his proper time. For as Nature has [illegible] bounds [illegible] na, [illegible] er- [illegible]

ACKNOWLEDGEMENTS

It is a pleasure to acknowledge the help from my colleagues in the preparation of this lecture. Dr. A. E. Kellie has very kindly supplied me with unpublished data on the urinary steroidal excretion, Dr. J. D. Robertson has given me the benefit of his wide experience for the section on basal metabolic rate, and I have to thank Dr. D. N. Baron also for his assistance.

REFERENCES

ALBRIGHT, F., SMITH, P. H., and RICHARDSON, A. M. (1941). *J. Amer. med. Ass.* **116**, 2465.
ANDERSON, I. A. (1949). *Quart. J. Med.* N.S. **73**, 67.
ANDRAL, G., and GAVARRET, J. (1843). *Ann. de Chir. et de Phys.* 3rd series. **8**, 129.
AUB, J. C., and DU BOIS, E. F. (1917). *Arch. intern. Med.* **19**, 823.
BACON, R. *The Care of Old Age and Preservation of Youth*, Trans. Richard Browne, 1683.
[illegible] 85, 610
[illegible] *l. Soc.* **71**, 143
[illegible] *Med.* **211**, 521
[illegible] *'s Hosp. Rep.* lxxi.
[illegible], J. D. (1941). *J. Amer. med. Ass.* **117**, 2144.
BURROWS, H. J., and GRAHAM, G. (1945). *Quart. J. Med.* N.S. **14**, 147.
CICERO, *De Senectute*. Loeb Classical Library.

COWDRY, E. V. (1942). Ed. *The Problems of Ageing*. Williams and Wilkins Co., Baltimore.

DAVIES, D. F., and SHOCK, N W. (1950). *J. clin. Invest.* **29**, 491.

DAVIES, D T., and JAMES, T. G. ILLTYD (1930). *Quart. J. Med* **23**, 93.

DOBRINER, K., GORDON, E , RHOADS, C. P., LIEBERMAN, S., and FIESER, L. F. (1942). *Science*, **95**, 534.

— LIEBERMAN, S , and RHOADS, C. P. (1948) *J. biol. Chem.* **172**, 241.

— LIEBERMAN, S., RHOADS, C. P., JONES, R. N , WILLIAMS, V. Z., and BARNES, R. B (1948). *J biol. Chem.* **172** 297.

— RHOADS, C. P., LIEBERMAN, S., HILL, B R., and SIESER, L F. (1944). *Science*, **99**, 494.

DUBLIN, L I. (1941) Statistical and Social Implications in the Problem of an Ageing Population, *The Diplomate*, **13**, 227.

DU BOIS, E. F. (1936). *Basal Metabolism in Health and Disease*, 3rd edn. Baillière, Tindall and Cox, London.

FLOYER, SIR JOHN (1725). *Medicina Gerocomica*, London.

[illegible]

— (1950) *Am J. Med.* 9, 229.

MAGNUS-LEVY, A , and FALK, E (1899). *Arch f Anat. u. Physiol. Supp. 315.*

— (1942). *J. Amer. med Ass* **118**, 1369.

MEULENGRACHT, E (1939). *Acta med scand.* **101**, 138.

SCHOENHEIMER, R. (1942) *The Dynamic State of Body Constituents*. Cambridge University Press

SHOCK, N. W (1951). *Trends in Gerontology* Stanford University Press, Stanford, California.

THEWLIS, M. W (1946) *The Care of the Aged* Kimpton, London.

XVI

Growth of the Human at the Time of Adolescence

J. M. TANNER

It is scarcely possible, in the space of a single hour, to convey very much information about a subject which is hardly dealt with at all in the medical textbooks, and which I must accordingly describe from the beginning. A good deal is known about the growth of boys and girls during their last years before maturity, and the knowledge was not come by easily. After the initial rough description of events through surveys of children of different ages—work which was for the most part satisfactorily completed before the end of the last century—further advances depended on following individual children throughout the whole course of their growth—the so-called longitudinal technique. Such a programme, demanded in particular by Franz Boas (1892), called for much perseverance in both scientific and administrative spheres, and for the ability to wait many years for results, with the knowledge that, if it eventually turned out that the wrong methods had been used, there was scarcely time for a second attempt.

Most of this hard-won solid core of knowledge concerns physical growth, of which we have now something approaching a complete picture. I shall summarize this as succinctly as I can. There is a very large literature on growth at adolescence; an admirable review by Harold Stuart (1946) covers it up to 1946, and in 1949 Shuttleworth (1949a, b) reissued two very valuable compilations of data, in the form of graphs and pictures. An attempt to pick out the prime contributors to our knowledge is a

difficult and dangerous undertaking, but I can scarcely discuss this subject without bringing to your attention the magnificent book on growth in general by D'Arcy Thompson (1942), B. T. Baldwin's (1921) classical monograph which has an unsurpassed bibliography of the older literature, the work of Franz Boas (1892, 1932, 1933, 1935), Richard Scammon (1930), Frank Shuttlewort
Palmer and
Krogman's
tions from
1948) such as the Fels Research Institute at Yellow Springs under Lester Sontag, the Child Research Council at Denver under Alfred Washburn, Harold Stuart's Center for Research in Child Health and Development at Harvard, the Brush Foundation study at Western Reserve, Howard Meredith's papers from the Iowa Child Welfare Research Station, and the University of California study, from which last year a particularly fine monograph was issued (Stolz and Stolz, 1951).

After outlining the main morphological events occurring about the time of adolescence, I am going to discuss for a moment their hereditary basis and then speculate a little upon their possible endocrinological causes. After that I shall move from theory to practice, and describe some of the organization and methods of the Harpenden Growth Study, with which I am associated and which is the first attempt in this country to study the growth of individual children from an early age to maturity. Lastly, returning to speculation, I shall endeavour to place the whole phenomenon of the human adolescent growth spurt in its biological background, by comparing the growth curves of man and other mammals.

PHYSICAL GROWTH AT ADOLESCENCE

In Fig. 1 is shown the growth curve of a single boy, measured every six months from birth to 17 years. Above the height of the child is plotted at successive ages; below the increments in height from one age to another. Growth is a form of motion; the upper curve is one of distance travelled, the lower of velocity. Velocity curves are as a rule more informative than distance

ones, particularly when it comes to relating morphological and physiological events, and for the rest of the time we shall rely upon them almost exclusively. This record is the oldest longitudinal study in existence, and it remains, for our purposes of

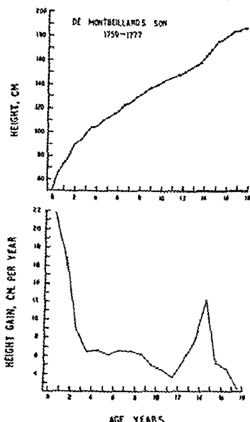


FIG. 1. Growth in height of de Montbeillard's son from birth to 18 years, 1759–77 Above, distance curve, height at each age below, velocity curve, increments in height from year to year. (Data in Scammon, 1927)

illustration, one of the best. It was made during the years 1759 to 1777 by Count Philibert Gueneau de Montbeillard upon his son, and was published by Buffon, a friend of Montbeillard's, in a supplement to his *Histoire Naturelle* (Scammon, 1927). You will see that in general the velocity of growth decreases from birth (and actually from the fourth month of foetal life) onwards,

but that this decrease is interrupted twice. It is checked between 6 and 8 years, a period known as the juvenile or mid-growth [illegible]

marked acceleration of growth which is known as the *adolescent growth spurt*.

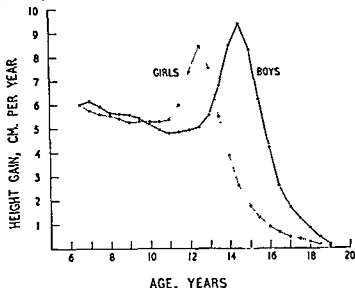
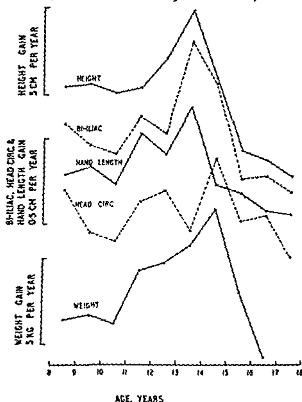


FIG 2. Adolescent spurt in height growth for girls and boys Each curve is from subjects who have their peak velocities between 12 and 13 for girls, and between 14 and 15 for boys (Redrawn from Shuttleworth, 1949)

The adolescent spurt is a constant phenomenon and occurs in all children, though it varies in intensity and duration from one child to another. In *boys* it occurs, on the average, from age 13 to 15½, and is responsible for a gain in height of about 8 inches (range 4–12 inches). The peak velocity of height growth averages about 4 inches (10 cm) per year, which is the rate the boy was growing at age 2 The time at which this maximum velocity is reached averages about 14 years, though it may lie anywhere between 12 and 17. In *girls* the spurt begins about 2 years earlier than in boys, lasts on the average from 11 to 13½, and is smaller

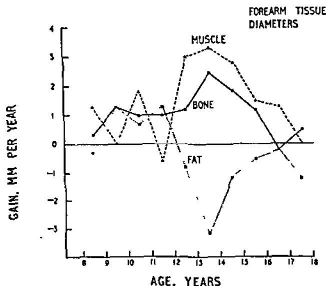
in magnitude, the peak height velocity averaging 3¼ in. (8 cm) per year. The sex difference can be seen in Fig. 2, redrawn from Shuttleworth (1939b), which shows the velocity curves for a group of boys who have their peak velocity between 12 and 13.



(These groups were chosen as being nearest the average time for peak velocity. Since their members were similar in time of the spurt the flattening and prolongation of the velocity curve which occur when data for people of different spurt times are averaged are avoided (Boas, 1892; Davenport, 1934; Shuttleworth, 1937; Tanner, 1951a).) The difference in size between adult men and women is to a large extent the result of this differ-

ence in adolescent spurt; prior to it, boys and girls are nearly the same size.

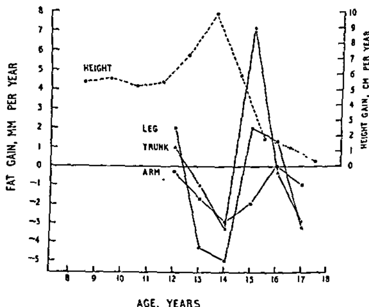
Every skeletal and muscular measurement in the body seems to take part in the spurt. Even the head diameters, dormant since a few years after birth, accelerate somewhat. The heart grows faster (Maresh, 1948; Bliss and Young, 1950), and so do



the viscera (Scammon, 1930; Castaldi and Vannucci, 1927). Fig. 3 shows, as an example, the velocity curves of stature, hip width, hand length, head circumference and body weight for a set of monovular triplets reported by Reynolds and Schoen (1947). The average of the three boys' figures are plotted for each measurement. There is a fairly regular order in which the various dimensions accelerate. leg length as a rule reaches its apex first, hip width follows a trifle ahead of shoulder width, with trunk length and chest depth the last of the skeletal dimensions. The relatively meagre data on muscular diameters sug-

gest they have their apex relatively late, as weight has also. The spurt in height is due more to an increase in the length of trunk than in the length of leg, and the ratio of trunk length/leg length is uniformly increased during adolescence.

To this regular progression of spurts only the *subcutaneous*



[illegible] Leg [illegible]

tissue is an exception. Subcutaneous fat decreases steadily in absolute amount from age one onwards until about a year before the height spurt begins. At that time, at least in boys, the fat increases, having a spurt of its own, seen by Meredith (1935) and Boynton (1936) but first described in detail by the Stolzs (1951). The fat spurt lasts about two years and then, when the general spurt gets going, changes to an actual loss of fat. This loss persists throughout the duration of the general spurt, but when the general spurt ceases, the fat comes back in a second wave which restores the amount to a level somewhat below that

reached at the peak of the original fat spurt. The fat spurts can thus perhaps most easily be regarded as a single wave interrupted and temporarily reversed by the general growth spurt. These changes are illustrated in Fig. 4 and Fig. 5, both on the triplet data. In Fig. 4 the measurements are of bone, muscle and fat diameters of the forearm as seen in a radiograph. The fat has its early adolescent spurt, then decreases as the bone spurts, followed by the muscle; the second increase in fat is not well shown in these data. In Fig. 5 measurements of fat of the arm, leg and trunk are plotted with stature (on a different scale) to give the time comparison.

RELATION OF BODILY GROWTH TO THE DEVELOPMENT OF THE REPRODUCTIVE SYSTEM

The adolescent spurt in skeletal and muscular dimensions is closely related to the spectacular development of the reproductive system which takes place at that time The sequence of events in the male is outlined in Fig. 6 (Schonfeld and Beabe, 1942; Greulich, Dorfman, Catchpole, Solomon and Culotta, 1942; Schonfeld, 1943; Stolz and Stolz, 1951; Reynolds and Wines, 1951). The solid areas marked *penis* and *testis* represent the period of accelerated growth of these organs, and the pubic hair ratings 0 to 6 follow the Stolzs' designation: (0) stands for no differentiated pubic hair, (1) unpigmented down 2–3 mm long, (2) slight pigmentation with sparse semi-terminals 4–10 mm long, (3) sparse terminals, pigmented, less than 25 mm long, (4) same at full density, (5) fully developed hair over 25 mm long, (6) extension laterally and/or upwards towards umbilicus. The sequence and timings given represent in each case the average value according to Stolz's data and so the entire picture is followed by few, if any, individuals The scheme represents boys in general in precisely the sense of Quetelet's 'homme moyen' or Viola's 'uomo medio'. To give an idea of the individual departures from this, figures for the range of ages at which the spurts for height, penis and testis growth begin and end are inserted underneath the first and last points of these curves. There are a few boys, it will be noticed, who do not begin their spurts in height or penis development until the

earliest maturers have quite completed theirs. At ages 13 and 14 there is an enormous variability amongst any group of boys, who range practically all the way from complete maturity to absolute pre-adolescence. The fact raises difficult social and educational problems and is itself a large contributory factor to the psychological maladjustments so often seen in adolescents.

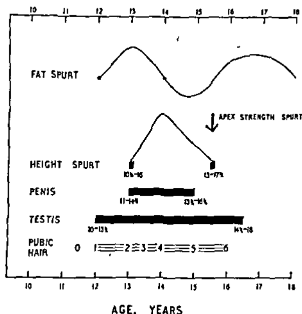


FIG. 6. Diagram of sequence of events at adolescence in boys. An average boy is represented the range of ages at which the events charted begin and end is given directly under the start and finish of each event.

In all cases, however, the testis spurt begins before the spurt of height or penis growth, and in all cases ends after the height and penis spurts have subsided. The usual first sign of impending puberty in the male is a beginning growth of the testes and scrotum. Slight growth of the pubic hair may begin about the same time, but proceeds slowly until the time of the general spurt, when it usually passes through stages 2, 3, 4 and 5 with some rapidity. Axillary hair begins to grow somewhat later than pubic hair, and body hair later still, with hair on the thigh, calf, abdomen and forearm usually preceding that on the chest and

upper arms. In individual boys there is a very close relation between timing of the height, testis and penis spurts so that the height spurt and penis growth begin rather regularly a year after the testis. The correlation coefficients for times of beginning and times of ending of these spurts in Stolzs' 67 boys are given in Table 1. The growth of the pubic hair, however, is not

TABLE 1. Correlation Coefficients between Ages at Beginning and Ending of Height, Penis and Testis Spurts. 67 Boys (Stolz and Stolz, 1951)

Beginning of height spurt and beginning of penis growth	0 87
Beginning of height spurt and beginning of testis growth	0 86
Beginning of penis growth and beginning of testis growth	0 85
Ending of height spurt and ending of penis growth	0 89
Ending of height spurt and ending of testis growth	0 89
Ending of penis growth and ending of testis growth	0 84

so closely bound up with skeletal and reproductive events, and varies a good deal in relation to them. There is a spurt in the development of muscular strength at adolescence, and its apex, which occurs fairly late, is shown in Fig. 6. If there is any validity at all in the conception that boys may 'outgrow their strength at adolescence', it applies to the couple of years, most often between 14 and 16, when maturity in height and reproductive organs is being rapidly achieved without the corresponding muscular spurt having yet fully developed.

A similar scheme for girls is shown in Fig. 7 (Preisel and Wagner, 1931; Shuttleworth, 1937, 1939; Simmons and Todd, 1938; Simmons, 1944; Reynolds and Wines, 1948). The rating of pubic hair corresponds to that for the boys, rating 1 not appearing since the authors on adolescence in girls have not noted this stage. Breast development stages follow Reynolds and Wines (1948). stage (1) is pre-adolescent, (2) the breast bud, elevation of breast and papilla as small mound, (3) elevation of breast and areola with no separation of their contours, (4) areola and papilla form a secondary mound above the level of the breast, (5) mature stage, papilla only projecting, due to recession of areola to general contour of the breast. The appearance of the breast bud is as a rule the first sign of puberty in the female, though the appearance of pubic hair may sometimes precede it. By analogy with the male, one would imagine that

[illegible] 3 and 14 t [illegible] of boys, [illegible] iturity to absolute pre-adolescence. The fact raises difficult social and educational problems and is itself a large contributory factor to the psychological maladjustments so often seen in adolescents.

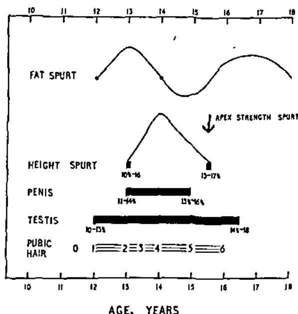


FIG. 6 Diagram of sequence of events at adolescence in boys. An average boy is represented; the range of ages at which the events charted begin and end is given directly under the start and finish of each event.

In all cases, however, the testis spurt begins before the spurt of height or penis growth, and in all cases ends after the height and penis spurts have subsided. The usual first sign of impending puberty in the male is a beginning growth of the testes and scrotum. Slight growth of the pubic hair may begin about the same time, but proceeds slowly until the time of the general spurt, when it usually passes through stages 2, 3, 4 and 5 with [illegible] than [illegible], calf, [illegible] t and

irregular than later ones, are anovulatory. Presumably similar considerations may apply to the male.

TABLE 2. Correlation Coefficients between Age at Menarche and [illegible]

Menarche and breast bud appearance	0 86
[illegible]	0 71
[illegible]	0 70
[illegible]	0 66

SEX DIFFERENCES ARISING AT ADOLESCENCE

The sex difference in body size is not the only one that arises at adolescence. Most of the differences between men and women in bodily shape, tissue composition, and physiological function arise then also. In the skeleton the chief distinction consists in men having relatively broader shoulders, narrower hips, and longer legs (see Tanner, 1951b). Fig. 8 shows how the shoulder-hip distinction arises. Girls have a particularly large spurt in hip width (Fig. 8 lower); a spurt, indeed, which is quantitatively about the same as that of boys, despite the boys' spurt being in all other dimensions so much the larger. The shoulder width spurt, on the other hand, is particularly marked in boys (Fig. 8 higher); the adult shoulder width depends more on the amount of growth at adolescence than do the other male skeletal dimensions The longer legs in the male come about through the slightly different mechanism of the spurt being later in boys; this is because in the immediately pre-adolescent years it is the legs which are growing relatively fastest of skeletal dimensions (as part of the general cephalad-caudad sequence of growth in the basic velocity curves) and thus if the body is allowed to grow for an extra two years before the spurt, the legs become relatively longer.

The chief part of the difference in shape between the male and female pelvis also arises at adolescence. Before puberty boys [illegible]

to the *acetabulum appears to push outwards* (Greulich and Thoms, 1944), chiefly due to a marked increase in growth of the

growth of the ovaries has already begun a year or so before this. The range of times for appearance of breast bud, stage (2) pubic hair and menarche are put in directly under the events in Fig. 7. The situation in regard to the early adolescent fat spurt is not clear, though there does seem to be a decrease in fat during the general spurt, and an increase after the general spurt is over

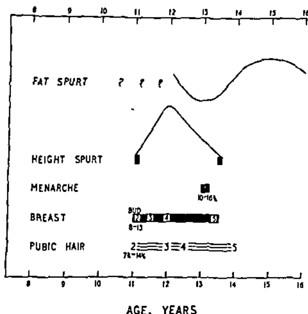


Fig. 7. Diagram of sequence of events at adolescence in girls. An average girl is represented: the range of ages at which some of the events occur is given directly under them

(Boynton, 1936; Reynolds and Grote, 1948). As in boys, the relation between height growth and reproductive organ development is closer than that of either with pubic hair growth The correlation coefficients between menarche and appearance of breast bud and pubic hair, and time of apex of height spurt are given in Table 2. Menarche occurs almost invariably after the apex of the height spurt has been passed. Full reproductive function does not appear till a year or more after menarche (Mills and Ogle, 1936; Ashley Montagu, 1946), and it is thought that the early menstrual cycles, which are more

irregular than later ones, are anovulatory. Presumably similar considerations may apply to the male.

TABLE 2. Correlation Coefficients between Age at Menarche and Age at Breast Budding, at Stage (2) Appearance of Pubic Hair (49 girls, Reynolds and Wines, 1948) and at Apex of Height Spurt (246 girls, Shuttleworth, 1937)

Menarche and breast bud appearance	0 86
[illegible]	0 71
[illegible]	0 70
[illegible]	0 66

SEX DIFFERENCES ARISING AT ADOLESCENCE

The sex difference in body size is not the only one that arises at adolescence. Most of the differences between men and women in bodily shape, tissue composition, and physiological function arise then also. In the skeleton the chief distinction consists in men having relatively broader shoulders, narrower hips, and longer legs (see Tanner, 1951b). Fig. 8 shows how the shoulder-hip distinction arises. Girls have a particularly large spurt in hip width (Fig. 8 lower); a spurt, indeed, which is quantitatively about the same as that of boys, despite the boys' spurt being in all other dimensions so much the larger The shoulder width spurt, on the other hand, is particularly marked in boys (Fig. 8 higher); the adult shoulder width depends more on the amount of growth at adolescence than do the other male skeletal dimensions. The longer legs in the male come about through the slightly different mechanism of the spurt being later in boys; this is because in the immediately pre-adolescent years it is the legs which are growing relatively fastest of skeletal dimensions (as part of the general cephalad-caudad sequence of growth in the basic velocity curves) and thus if the body is allowed to grow for an extra two years before the spurt, the legs become relatively longer.

The chief part of the difference in shape between the male and female pelvis also arises at adolescence. Before puberty boys have smaller sciatic notches than girls, but there is no distinction in the shape of the inlet (Reynolds, 1945, 1947; Washburn, 1948). At puberty the portion of the pelvis immediately medial to the acetabulum appears to push outwards (Greulich and Thoms, 1944), chiefly due to a marked increase in growth of the

pubic bone and the lower part of the ilium (Washburn, 1948). Probably this is due to a great development of the ossification centres which appear shortly before puberty in the Y-shaped cartilage joining the ilium, ischium and pubis across the acetabulum. The effect is almost certainly due to direct stimulation

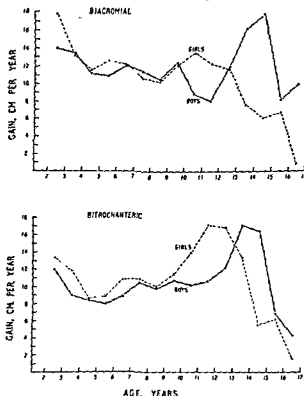


Fig 8 Velocity curves of shoulder width (biacromial) and hip width (bitrochanteric) in girls and boys. Mixed longitudinal. (Data from Simmonds, 1944)

by oestrogen at this time; the pelvis of the hypogonadal man differs not at all from that of the normal male (Greulich and Thoms, 1939). Probably the male growth of the shoulders similarly represents a direct response to androgens.

Our chief knowledge of tissue composition relates to the calf, because of the ease with which bone, muscle, and fat can be distinguished there in radiographs (Stuart and Dwinell, 1942;

Reynolds, 1944; Stuart and Sobel, 1946; Reynolds, 1946; Reynolds and Grote, 1948; Reynolds, 1949; Stuart and Reed, 1951). There is a characteristic sex difference in the adult calf; so much [illegible] there [illegible] cent [illegible], and women more fat, and this difference arises almost entirely at adolescence; differentiation of sex by calf X-ray at age 7½ gives results very little better than chance guessing.

PHYSIOLOGICAL CHANGES AT ADOLESCENCE

The chief studies of physiological changes at adolescence, apart from the endocrinological ones we shall discuss later, have been made by Nathan Shock (1941, 1942, 1943, 1944, 1946a). Fig. 9, redrawn from one of his papers, shows the basal systolic blood pressure, heart rate, and oxygen consumption of fifty girls studied every six months. The curves are drawn in relation to the time of menarche, since the physiological changes of adolescence, as one would expect, are more closely related to the growth spurt and to maturation of reproductive function than to chronological age. If the spurt occurs early, the physiological events occur early too; if the spurt occurs late, so do they. In Fig. 9 the beginning of the general growth spurt occurs about one or two places from the left-hand margin of the curves. It will be seen that heart rate rises during the spurt and falls after the spurt finishes, and oxygen consumption does likewise. Systolic blood pressure (Richey, 1931; Downing, 1947) rises to its adult values during the spurt, and for boys rises further than for girls, so that the adult basal difference between men and women is another characteristic established at adolescence (cf. biacromial diameter, Fig. 8) Probably this is due to the establishment of a greater basal stroke volume of the heart at this time (Nylin, 1935), coincident with the larger heart size of males. Between diastolic pressures there is no difference. At the same time in boys the number of red blood cells rises and consequently the haemoglobin increases, without change in mean corpuscular volume, haemoglobin, or haemoglobin concentration (Mugrage and Anderson, 1938). Girls lack this rise, which establishes the

adult sex difference. The physiological responses to exercise also change (Shock, 1946b) and the basal alveolar CO_2 pressure rises in boys, though not in girls. Other studies confirm the rise in basal metabolic rate reported as oxygen consumed per square metre surface area seen in Fig. 9 (Topper and Mulier, 1932;

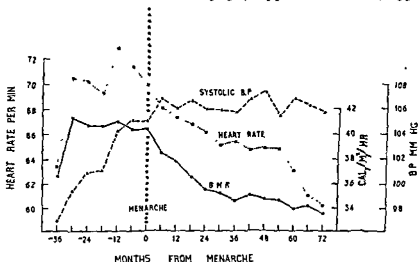


Fig. 9. Physiological changes at adolescence in relation to menarche. Adolescent height spurt runs from about − 24 months to + 6 months. (Redrawn from Shock, 1946.)

Bruen, 1933; Nylin, 1935; Talbot, Wilson and Worcester, 1937) or show at least a marked check at adolescence in the previously steady rate of fall (Lewis, Duval and Iliff, 1943a, b). As muscle is put on the excretion of creatinine/body weight rises (Sandford, Adkins, Miller and Cobb, 1943), more so in boys than in girls (Clark, Thompson, Beck and Jackson, 1951), and the excretion of creatine diminishes or ceases (Shock, 1945). Changes in strength and in muscular performance also occur at about this time (Dimock, 1935; Espenschade, 1940, Jones, 1949).

THE RELATION OF THE TIME AND CHARACTER OF THE ADOLESCENT SPURT TO ADULT PHYSIQUE

There are differences in physique between those who mature early and those who mature late and these differences can be seen both before adolescence has at all started, and after it is

entirely over. (During adolescence, of course, these differences are obscured or magnified by the early maturers being larger than the late maturers because of the adolescent spurt itself.) As long ago as 1921 Baldwin came to the conclusion that boys and girls who were tall before puberty began adolescence earlier than those who were short. Meredith (1935), Dimock (1935) and Boynton (1936) briefly confirmed this finding, but the first [illegible] age 13, from 13 to 14, or after 14. In Fig. 10, redrawn from his data, are given the distance curves (i.e. actual amount reached at each birthday) for height and weight from age 6 to 17. The early spurting girls are greater in height and weight at ages 6 and 7 and 8, before the influence of the spurt can make itself felt; they are large girls before puberty and the effect for both height and weight is similar. At maturity, however (age 17), there is no difference in height between the three groups. some other data actually show the late maturers a trifle taller (Stone and Barker, 1937). The weight difference, however, persists. The same holds good for boys. Early maturers are firstly people whose growth to maturity is more advanced *at all ages* than late maturers, and are secondly people who are less linear or ectomorphic (Pryor, 1936; Bayley, 1943b; Simmons and Greulich, 1943; Reynolds, 1946; Reynolds and Wines, 1948).

Besides, or as a corollary to this, Bayley (1943b) has adduced evidence that there are differences in androgyny—the degree of masculinity in the female or femininity in the male—between adults who experience an early or late puberty. According to her data, early maturing boys have, as adults of 18, relatively shorter legs and broader hips than late maturers. The pattern of their spurt approximates more nearly to that of girls, in composition as well as in time, and their final build does so likewise. Late maturing boys have, on the other hand, long legs and narrow hips. Conversely, late maturing girls approximate somewhat to the male pattern and at maturity have relatively long legs and broad shoulders. This study still needs confirmation on a more extensive series of subjects, and examination of the

Stolzs' data for boys does not appear to give the same results. However, some of the late maturing boys of this latter study appear not to have finished growing when last measured, so that it is not really possible to judge the matter from these figures.

The Stolzs have themselves described a somewhat different relation between type of spurt at adolescence and adult physique. Those boys whose early adolescent fat spurt was marked

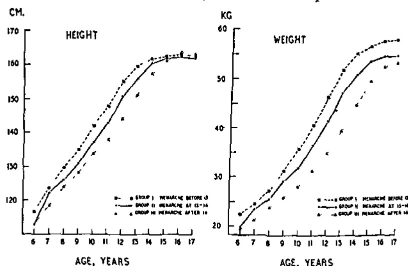


FIG 10 Growth in height and weight of early, average, and late maturing girls Mixed longitudinal data. (Redrawn from Richey, 1937)

—and they are both early and late maturers—when fully grown were more androgynous than those whose fat spurt was slight. During their spurt, these boys experienced considerable breast development, with increase of the nipples and areola They also had a more feminine type of thigh-leg configuration, that is, knock-knees rather than bow-legs (see Draper, Dupertuis and Caughey, 1944; Seltzer, 1945; Bayley and Bayer, 1946) and a low shoulder width/hip width ratio, and they retained most of these characteristics after maturity was reached. They did not have a smaller penis than other boys, however, nor any less pubic or body hair

Despite what I have said, I do not want to leave the impression that the adolescent spurt causes any radical change in body

build: it certainly does not. It adds only the finishing touches to a physique which is recognizable years before. Anyone who has looked at serial pictures of children followed from infancy to adulthood must be impressed chiefly by the similarity the child shows from one age to another. So great is this that a man used

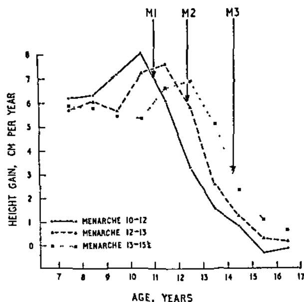


FIG. 11. Relation of peak velocities in height for early, average, and late maturing girls, and to show time elapsing between peak velocity and menarche for the three groups M1, M2, M3, average time of menarche for each group. (Redrawn from Simmonds and Todd, 1943)

to looking at children's photographs could probably predict the adult somatotype with accuracy from a picture taken at age 5 or even earlier.

Two points remain to be made in connexion with the timing of the adolescent spurt. The first is that the earlier the spurt, the more intense it is, a fact discovered by Boas (1932) and thoroughly investigated by Shuttleworth (1937). This is well exemplified in the data of Simmons and Greulich (1943), from which Fig. 11 is drawn. Not only is a greater peak velocity of height reached by the early maturers, but the whole process appears to

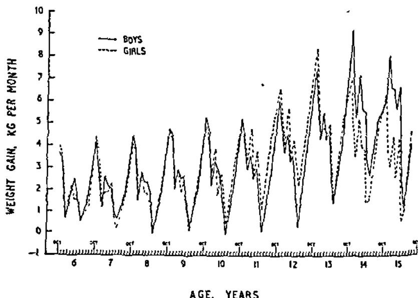


Fig. 12. Seasonal variation of weight increments in boys and girls of U.S P.H.S. Hagerstown Survey, 1921–8 (Redrawn from Palmer, 1933)

be going more quickly in them, and the time from peak velocity to menarche is shorter in them than in the others.

Secondly, there is a well-marked *seasonal effect* on velocity of growth in the human (Fitt, 1941), with growth in height greatest in spring and growth in weight greatest in autumn. Adolescence provides no exception to this, so that the velocity of weight growth varies throughout the year in the way illustrated in Fig. 12, taken from Palmer (1933).

FACTORS CONTROLLING TIME AND COMPOSITION OF THE ADOLESCENT SPURT

A certain amount is known about factors affecting the time at which the adolescent spurt begins. Fundamentally, the control is a genetic one, and there is a fairly high correlation between the times of menarche of mothers and daughters (Bolk, 1923; Popenoe, 1928; Gould and Gould, 1932), and a higher one between twins (Verschuer, 1927, Petri, 1934) This control operates throughout the whole process of growth, the rate of development being evidently dependent on a number, and pro-

bably a fairly large number, of genes (see Reynolds, 1943; Tanner, 1953). The adolescent spurt begins when a certain degree of maturity has been reached, or when a certain amount of the growth process has been gone through. The requisite maturity is related to some extent, but not too closely, to chronological age, and to some extent to size achieved (Palmer and Reed, 1935; Palmer, Kawakami and Reed, 1937). It is more nearly approximated by the degree of maturation of the ossification centres of the hand, wrist and knee (the skeletal age; Florey, 1936; Shuttleworth, 1938; Simmons and Greulich, 1943, Bayley, 1943a; Greulich and Pyle, 1950), and perhaps by the number and degree of eruption of the teeth (Boas, 1932; Shuttleworth, 1937) though this latter is uncertain.

Girls are ahead of boys in maturity, judged by skeletal ossification, even at birth (Menees and Holly, 1932; Kelly and Reynolds, 1947; Christie, 1949), and continue to be so by this criterion and by the criterion of permanent tooth development right up to adolescence. (But not in eruption of deciduous teeth apparently (Meredith, 1946).) Admittedly girls are slightly smaller than boys at birth and up to adolescence, and presumably they would be slightly smaller still if they were not ahead in development. It seems that it is the brain, and probably the hypothalamus, which initiates the events of the adolescent spurt (see below) and how the brain can tell when the requisite pre-adolescent development is complete we have no idea. One thinks of a series of clocks, the running-down of one being the signal for the starting of another, but the clocks in the brain are highly dependent on the growth and development of the body and very little dependent on chronological time. Starvation in animals or in man postpones the time of adolescence (Backman, 1948, Youmans, 1948), and it seems that in rats adolescence under these circumstances simply waits until the body has grown, however slowly, to approximately its normal adolescent size, or, more probably, to its requisite maturity.

Since girls precede boys throughout development, it seems likely that speed of maturation, like sex itself, depends on the quantitative balance between genes on the X chromosomes and genes on the autosomes. On this view boys with early adolescent

spurts would have an autosome-sex chromosome balance nearer to that of girls than would late maturing boys. It has recently been shown that on the X chromosome of the human there are genes increasing body breadth but no genes increasing stature (Tanner, unpublished). This seems to imply that body breadth and time of adolescence are partially sex-linked characters, with at least a part of these two effects probably due to the same genes, whereas the sex difference in stature is sex-limited, and dependent on the androgen secretion at adolescence.

THE ENDOCRINOLOGY OF ADOLESCENCE

It seems certain that the events of the adolescent growth spurt take place under hormonal control. To account for the various phases of adolescence—the early fat spurt, then the general growth spurt, the growth of pubic hair, and so forth—we must clearly postulate a fairly complex series of changes either in secretion rates of various hormones, or in the responsiveness of various tissues to the hormones or, possibly, in both. Our data on this are exceedingly limited—there seem to be only two longitudinal studies of hormone excretions and they only cover a couple of years—and what I have to say now must be regarded as somewhat speculative. So far as I am aware there exists at present no coherent theory of the endocrinology of adolescence; the discussion in Wilkins' (1950) text and the paper by Kinsell, Michaels, Li and Larsen (1948) perhaps come nearest to one. But if due attention is given to the sequence of observed morphological changes, it is at least possible to indicate a set of reasonable alternatives.

The endocrinological data we have to go on consist of twenty-four-hour [illegible] by bioassay, androger[illegible] reducing lipids by [illegible] sectional, so that the patterned individual changes we are particularly concerned with are inevitably smoothed over and missed. However, such as they are, the excretion studies can be summarized thus:

(a) *Gonadotrophins* are scarcely if at all detectable before the beginning of adolescence, but appear at about the time the testis

begins to enlarge in boys, and at a similar time in girls. Their excretion rises fairly rapidly during adolescence, reaching in boys the maximal level of 6 rat units at maturity, and in girls the higher value of 10–15 rat units. Greulich and his associates (1942) have reported excretions in boys in relation to the boys' maturity group as judged by his secondary sex characters. The difficulty of early and late maturers is thus removed. These results are given in Table 3; maturity group (1) is prepubescent,

TABLE 3. Urinary Excretion per twenty-four hours of Androgens, Oestrogens and Gonadotrophins of Normal Boys according to Secondary Sex Character Development (Greulich *et al.*, 1942)
For details see text.

Maturity group	1	2	3	4	5
Androgen, I.U. androsterone	4 1	6 6	10 0	19 7	28 8
Oestrogen I.U. oestrone	7 0	11 4	16 7	37 8	45 9
Gonadotrophin M.U.U.	<3	<3	ca. 3	ca. 5	ca. 7

(2) signifies beginning penis and testis growth with pubic hair stage 2, and corresponds to a point about opposite age 13 in Fig. 6, (3) represents a point opposite about age 14 in Fig. 6, (4) a point opposite about $14\frac{3}{4}$, and (5) opposite $15\frac{1}{2}$. In this study gonadotrophin bioassay was by the increase in uterine weight in immature mice. Pedersen-Bjerrgaard and Tønnesen (1948a, b) also used the uterine weight increase in immature mice or rats, and found that girls from 3 to 12 excreted 1 or less than 1 rat unit of gonadotrophin per 24 hours, while from 12 to 16 excretion rose to 15 rat units. It seems that there was a peak excretion at this time, with a fall to a value of 9 rat units at about 20, which is then maintained till the mid-thirties. However, the peak may be a chance sampling effect; the data are scarcely sufficient to prove otherwise. In boys gonadotrophin excretion became detectable at about 8 years old and rose to a value of 6 rat units at 15, and continued at this level until the sixth decade. Nathanson, Towne and Aub (1941) used stimulation of follicle growth in the ovary of immature mice as the criterion of gonadotrophic excretion and could detect none before the age of 13 in boys and 11 in girls.

(*b*) *Oestrogen* excretions follow the course shown in Fig. 13, and in Table 3. There is a very low and roughly constant excretion of oestrogenic substances by both boys and girls from 3 to 7. After about 7 the excretion gradually rises, equally still in both

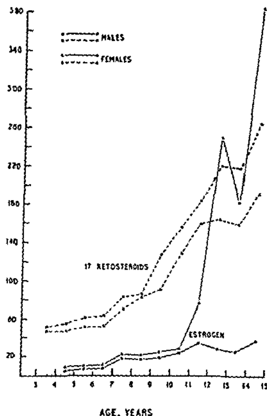


Fig. 13 Excretion of oestrogen and neutral 17-ketosteroids by girls and boys. Vertical scale is I.U. oestrone, and colour units each equivalent to 60 μg androsterone (Redrawn from Nathanson, Towne, and Aub, 1941)

sexes until adolescence supervenes in the girls. At this time the excretion in girls rises very sharply and begins to be cyclic with even the minimum monthly value considerably above that for boys. Oestrogen cycles are established, it seems, at about the time the breast buds appear and the spurt in height begins: the values reached in the cycles increase until some years after

the menarche (Pedersen-Bjerrgaard and Tønnesen, 1948a). The three studies have all used the same bioassay technique, the appearance of cornified cells in the vaginal smear of spayed mice.

(*c*) *Androgenic substances* (by bioassay on comb growth in the day-old chick or capon) begin to become detectable in the urine of boys and girls at about 8 years old, and their excretion increases gradually until adolescence, when a considerable rise takes place in both sexes: girls increase to an average twenty-four-hour excretion of about 20 I.U. androsterone, boys to about 30 I.U. at 18. Boys continue to increase to 40 or 50 I.U. in the mid-twenties (Hamburger, Halvorsen and Pedersen, 1945; Pedersen-Bjerrgaard and Tønnesen, 1948b). The results of Greulich and his associates are given in Table 3.

(*d*) The excretion of *17-ketosteroids* is shown in Fig. 13. It is very similar to that for androgens, with a gradually increasing level from 7 or 8 onwards, the increase becoming sharper in both sexes at puberty. No sex difference is visible until the time of the boys' adolescence, when the adult difference is established, men having a greater excretion than women (Oesting and Webster, 1938; Nathanson, Towne and Aub, 1941; Hamburger, 1948; Pedersen-Bjerrgaard and Tønnesen, 1948b; Wood and Gray, 1949).

(*e*) *Neutral reducing lipids* (or cortisone-like hormones) are excreted at a gradually increasing rate from birth to maturity, equally in both sexes. The increase is proportionate to the increasing body size and no particular spurt seems to occur at puberty, at least in the only data at present available (Talbot, Wood, Worcester, Christo, Campbell and Zyguntowicz, 1951). Data on the excretion of glycogenic corticoids (deposition of glycogen in the liver of adrenalectomized mice) have been given by Venning and Kazmin (1946) and are in accord with this.

The interpretation of these facts so far seems to be fairly straightforward. The major and possibly only source of oestrogens in the male and androgens in the female is the adrenal cortex and the conclusion seems inescapable that at adolescence the cortex fairly rapidly increases its secretion of sex hormones, just as do the gonads. This adrenal component of adolescence is

sometimes rather inelegantly as well as inaccurately referred to as the 'adrenarche'. It is under the control of the pi-
it is absent
(
i ... excretion argues rather strongly against ACTH as the stimulus, and in favour of the view that some separate pituitary hor-
and perhaps also
gonadotrophin, an ... particular luteinizing or interstitial-cell-stimulating hormone, has been mooted. Such a view, championed especially by Fuller Albright, would certainly fit the facts of hormone excretion in children very nicely. (This speculation could be pushed further. Oestrogen stimulates the secretion of LH by the pituitary, and it might be that FSH first acted upon some cells in the adrenal causing them to secrete estrogen, which stimulated the secretion of LH which in turn acted upon other cells in the adrenal causing them to produce androgen. Adrenal androgen secretion ...
not dependent
increase in adr... oestrogen secretion seems much less prominent than the androgen increase, but occurs nevertheless, and in both sexes; the oestrogen output in a series of young adult ovariectomized women is given by Pedersen-Bjerrgaard and Tonnesen (1948a) as 15 per cent of the average of normal women. Lastly it should be remembered that practically all this hormone ...
c
n

a
se ... hormone levels, the interpretation of none of these is clear yet, however. Clark and Beck (1950) give values for *plasma alkaline phosphatase* which, in concert with the results they quote from previous studies, show a marked parallelism between plasma level and velocity of growth in height. An adolescent rise is clearly seen in boys and is probably also present in girls, to a lesser degree, and as bone growth ceases, so the phosphatase drops to adult levels. Since it has been shown

that injection of pure growth hormone raises plasma alkaline phosphatase and hypophysectomy lowers it (Li, Kalman, Evans and Simpson, 1946), it is tempting to suppose that these levels in children reflect growth hormone secretion. This is as yet stretching the evidence too far, however, since there are other factors known to raise plasma phosphatase in animals, in particular testosterone, which may be responsible for the male rise at adolescence. Probably alkaline phosphatase reflects the level of bone growth irrespective of the growth stimulus. *Serum inorganic phosphorus* level is also affected by growth hormone, and is high in acromegaly (which alkaline phosphatase, the epiphyses being closed, is not). There seems to be only one thorough study of the level of this substance in childhood, by Stearns and Warweg (1933). The data are cross-sectional and indicate that either no change or else a very slight but steady fall takes place between 1 and 15 years. (Phosphatase shows a continuous and considerable fall.) At about adolescence—the data are not of the sort that show for certain whether before, during or after the growth spurt, but it appears to me probably as the spurt finishes—the level falls to adult values. If this interpretation of the data is correct, it points towards a decreased growth hormone secretion after cessation of the adolescent spurt, with a persisting secretion during it. Kinsell, Michaels, Li and Larsen (1948) have argued otherwise, but not, to my mind, at all convincingly.

Lastly the excretion of creatine and creatinine has been considerably studied (Nathanson, Miller, Towne and Aub, 1941; Shock, 1945; Reynolds and Clark, 1947; Clark, Thompson, Beck and Jacobson, 1951). Creatinine excretion has a curve very similar to that for 17-ketosteroids. There is an increasing rise from 5 to about 12, with boys and girls practically indistinguishable, and thereafter a greater rise in the boys' excretion. It is not clear however whether the sex difference signifies primarily testosterone secretion or merely accompanies the increase of muscle mass.

It is on such meagre data that we must construct some theory of endocrinological relationships during adolescence. Taking the male spurt, we may suppose that at a certain level of maturity, usually about age 12, gonadotrophins begin to be secreted,

probably for the first time, and cause enlargement of the testis and scrotum, and after a delay of a year or less the secretion of testosterone by the testis. The testosterone causes the penis to grow, and probably all or most of the various physiological changes which occur to differentiate boys and girls at this time. At the same time as the gonadotrophins are first released, something is happening to cause the early adolescent fat spurt, and for this there seem to be two possibilities. Both oestrogen (Nyda, de Majo and Lewis, 1948) and cortisone cause deposition of subcutaneous fat. The excretion evidence is against cortisone, and in favour of oestrogen, and, much more convincingly, many boys experience definite breast and areolar effects, which can scarcely be attributed to anything but oestrogenic stimulation. Also those boys with marked fat spurts acquire, a year or so later, large hips and a feminine configuration of the legs which are almost certainly due to oestrogen. (Presumably such an occurrence does not necessarily imply decreased testosterone or adrenal cortex secretion, since the penis size and pubic hair of these boys are not below average.)

In young adults, on the other hand, the amount of subcutaneous fat is correlated with the level of serum cholesterol (Tanner, 1951c), and this relationship probably comes from cortisone action on both; rather than oestrogen action, since there is no difference in serum cholesterol, or in cortisone excretion (Talbot *et al.*, 1951) between men and women. The adult level of serum cholesterol seems probably to be established by two months of age (Hodges, Sperry and Andersen, 1943), a result consonant with the childhood excretion of cortisone. It may well be—though this is nearly pure speculation—that the differences in subcutaneous fat not allied to sex and present from birth onwards are related to cortisone secretion (plus other [illegible]

Involution of the thymus occurs at about the time of adolescence, though unfortunately the data (Boyd, 1936) are not sufficient to show the relations of this to the other events. Since both cortisone and sex hormones (Korenchevsky, Dennison and

Simpson, 1935; Reinhardt and Wainman, 1942; Feldman, 1951) cause involution experimentally, the occurrence does not as yet help forward our interpretations.

About a year after release of gonadotrophins, when the fat spurt is maximal, the general growth spurt beings. Again there seem to be two endocrinological possibilities. Testosterone from the testis certainly accounts for some or all of the difference between male and female spurt, but it cannot account for the existence of the spurt in girls. The chief element in the spurt must be either increased secretion of growth hormone or else the increased androgen secretion from the adrenal: oestrogen is ruled out as it does not cause growth of bone or muscle (Johnston, 1941). The biochemical pattern of growth at this time consists in laying down of protein and utilization of fat for energy purposes with loss of fat from the body; and this is the typical effect pro-
[illegible] Pro-
[illegible] pic-
[illegible] 950)
which may imply a stimulation of growth hormone activity at this time. In favour of the adrenal hypothesis is the undoubted increase in excretion of adrenal hormones, and perhaps the recent demonstration that several steroids exist (such as methyl-androstenediol) which in general resemble testosterone but have greater anabolic effects on muscle and possibly bone and smaller effects on the reproductive organs. It is certainly possible that steroid-anabolic hormones such as these could be the cause of the female adolescent spurt and that portion of the male spurt not directly attributable to testosterone. In favour of this are the cases of premature pubic hair growth in girls, usually accompanied by growth in height and weight, but without evidence of virilization (Wilkins, 1950).

The relatively greater growth of bone width in the male calf at puberty, however, argues in favour of this effect being primarily a testosterone one, just as the fat growth in the female calf is presumably due to oestrogen.

Apparently a phenomenon analogous to the early adolescent fat spurt in boys occurs in girls. Some girls in late adolescence become rather more muscular than the majority, grow some-

what more body hair, and may have a slight growth of the clitoris (Wilkins, 1950). These are evidently the girls who are high in masculinity, just as the early-fat-spurt boys are the boys high in femininity. Probably these girls also have above-average growth of the shoulders; no specific study of them has been made. Presumably they secrete more adrenal androgen than the others, or else secrete androgens which have somewhat more nearly the action of testosterone.

The spurt ends with the closure of the epiphyses, and this was at one time attributed to the direct action of testosterone or oestrogen. So far at least as testosterone is concerned the evidence is entirely against such a view. Testosterone administered to prepubescent boys (or girls) to cause increased growth does not at the same time cause premature closure of the epiphyses (Deamer, 1948; Wilkins, 1950; Escamilla and Bennett, 1951). Probably what causes the epiphyses to unite is the cessation of growth hormone secretion. At least this applies to the epiphyses of the long bones; the secondary epiphyses of the vertebrae and pelvis and shoulder region seem more under the control of steroids. Giants, no less than eunuchs, have long arms and legs, and they go on growing till they are 25 or 30, and their sex development is often normal (Wilkins, 1950). Growth hormone is probably secreted from the age of 4 or so onwards to adolescence, when it may or may not temporarily increase. It then gradually ceases, probably due to inhibition by the increasing oestrogen or testosterone level; in eunuchs no testosterone is secreted, and growth hormone production ceases later than usual, giving them the characteristically long arms and legs.

In this carefully patterned hormonal whirlpool we can make very little attempt to place insulin, though we may suppose it to have quite an important rôle About its relations with growth hormone in man too much is obscure even for reasonable guesswork, and there are no studies of insulin action during this period.

Lastly, as to the cause of hair growth on the pubes and elsewhere. It seems that androgens from the testis and adrenal both stimulate hair growth and possibly oestrogens do so as well. The differential growth at the pubes, axilla and face seems most

easily explicable on the basis of locally different thresholds to stimulation, coupled with a predilection of hair at each site for either testis, ovarian or adrenal hormone. On this hypothesis, the skin of the pubes has the lowest threshold and responds to the small amount of androgen secreted by the adrenal in both men and women early in puberty, at which time the hair at this spot begins to grow. We have already noted that the development of pubic hair takes place rather independently of the reproductive tract, and this adrenal control is presumably the reason why it does so. In ovarian agenesis there is usually a sparse growth of pubic hair, though sometimes none at all, and sometimes a normal amount. Treatment with oestrogen in this condition causes the pubic hair to grow (Lisser, Curtis, Escamilla and Goldberg, 1947). This may be a direct effect or possibly a secondary one from the adrenals following pituitary LH release caused by the oestrogen. The axillary hair has a somewhat higher threshold, develops later, and is possibly more responsive to testosterone: it develops only slightly in male castrates (Hamilton, 1951). The beard has a still higher threshold to adrenal androgens, and a more pronounced preference for testosterone. The general body hair seems to have a higher threshold, but to be affected by both adrenal and testis. The relative profusion of body hair seems to run in families (Greulich *et al.*, 1942; Reynolds, 1951), and probably somewhat independently of sex; this may be due either to differences in skin responsiveness or in amount of androgens secreted.

Possibly this complicated sequence of endocrinological events is self-controlling once started, but its initiation, it is quite clear, is under the control of the brain Gonadotrophins, the secretion of which starts the whole mechanism going, are present in animal pituitary glands long before puberty (Severinghaus, 1942), but are not released to the circulation Recently Harris and Jacobson (1951) have shown that when the pituitary of a normal rat is grafted into a hypophysectomized animal under [illegible]

newborn pituitary secretes gonadotrophin as soon as vasculariza-

z

tion is complete and long before the donor rat would have reached puberty. This is very convincing evidence that it is the hypothalamus which has to mature before adolescence begins, not the pituitary. The same is shown in the human by the literature on precocious puberty of cerebral origin. Ganglioneuromas of the tuber region may cause all the changes of adolescence to take place as early as 2 years old (Le Marquand and Russell, 1934–5; Gesell, Thoms, Hartman and Thompson, 1939; Weinberger and Grant, 1941; Troland and Brown, 1948; Seckel, Scott and Bendit, 1949; Seckel, 1950). Some authors also think that tumours of the *posterior* hypothalamus and of the epithalamic region may do the same, but this is less well established. Certainly precocious puberty may on occasion follow encephalitis. It appears also that precocious puberty may occur hereditarily in otherwise perfectly healthy persons (Rush, Bilderback, Slocum and Rogers, 1937; Engstrom and Munson, 1951), presumably due to a disturbance of the maturity mechanism by an abnormal hypothalamus. Such precociously occurring puberties may be in no way different from the ordinary event so far as reproductive function is concerned. The boy may progress to active spermatogenesis and the girl to ovulation, corpora lutea and, perhaps, pregnancy (Wilkins, 1950). Mental development, development of the teeth, and psychosexual advancement do not, however, keep pace. it seems in regard to the last of these that the hormones need a mature brain to work upon.

The facts of precocious puberty argue rather strongly against the occurrence of any marked increase of sensitivity of tissues being responsible for the hormonal effects of puberty, unless the changes themselves are hormonally conditioned, rather than caused by the tissues in some way maturing. The overriding control of the pituitary by the hypothalamus should not be taken to mean that the pituitary does not itself mature, at least in the normal. The anterior pituitary, on the contrary, has a spurt in weight growth at adolescence in both sexes, and a sex difference appears with the female gland in this instance growing larger than the male and acquiring a higher percentage of acidophils (Rasmussen, 1947, 1950). In rats both increased

growth of the pituitary and increased numbers of acidophils can be brought about by injection of oestrogen (Baker and Everett, 1947), and can be prevented from occurring in normal females by administration of testosterone (Wolfe, Wilson and Hamilton, 1944).

THE HARPENDEN GROWTH STUDY

Before we proceed to the evolutionary aspects of the adolescent spurt, let me interpose a short description of the Harpenden Growth Study, to give you some idea of the practical background that the results I have described spring from.

The Harpenden Growth Study owes its beginning to the initiative of Dr. E. R. Bransby of the Ministry of Health Scientific Staff. During and after the war he had been concerned with various surveys of child growth, mostly conceived with an eye on nutritional status and the rationing system. As a result of this Dr. Bransby came to the conclusion that nothing less than the establishment of a full-fledged and more or less permanent longitudinal study of children would suffice to answer the practical questions he was concerned with, in particular the usefulness of growth records in the School Medical Service. I had for some time been advocating the same from the point of view of the scientific investigation of human growth and so he contacted me towards the end of 1947 and we made plans together for setting up a comprehensive longitudinal study. The following summer I was enabled through the generosity of the Wenner-Gren Foundation to spend some time at each of the major longitudinal studies in America, so that we were able to begin with at least a small amount of that first-hand technical experience which is irreplaceable and invaluable in setting up projects of this kind.

The Study, for which the Ministry of Health and the Sherrington School of Physiology of St. Thomas's Hospital are jointly responsible, was sited at the Harpenden Branch of the National Children's Home, because previous experience with the staff and children there had shown us how exceedingly helpful and friendly they were. Nor were we disappointed in making our larger demands on their time and patience; the Study owes its entire existence to the efforts on its behalf of the

tion is complete and long before the donor rat would have reached puberty. This is very convincing evidence that it is the hypothalamus which has to mature before adolescence begins, not the pituitary. The same is shown in the human by the literature on precocious puberty of cerebral origin. Ganglioneuromas of the tuber region may cause all the changes of adolescence to take place as early as 2 years old (Le Marquand and Russell, 1934–5; Gesell, Thoms, Hartman and Thompson, 1939; Weinberger and Grant, 1941; Troland and Brown, 1948; Seckel, Scott and Bendit, 1949; Seckel, 1950). Some authors also think that tumours of the posterior hypothalamus and of the epithalamic region may do the same, but this is less well established. Certainly precocious puberty may on occasion follow encephalitis. It appears also that precocious puberty may occur hereditarily in otherwise perfectly healthy persons (Rush, Bilderback, Slocum and Rogers, 1937; Engstrom and Munson, 1951), presumably due to a disturbance of the maturity mechanism by an abnormal hypothalamus. Such precociously occurring puberties may be in no way different from the ordinary event so far as reproductive function is concerned. The boy may progress to active spermatogenesis and the girl to ovulation, corpora lutea and, perhaps, pregnancy (Wilkins, 1950). Mental development, development of the teeth, and psychosexual advancement do not, however, keep pace. it seems in regard to the last of these that the hormones need a mature brain to work upon.

The facts of precocious puberty argue rather strongly against the occurrence of any marked increase of sensitivity of tissues being responsible for the hormonal effects of puberty, unless the changes themselves are hormonally conditioned, rather than caused by the tissues in some way maturing. The overriding control of the pituitary by the hypothalamus should not be taken to mean that the pituitary does not itself mature, at least in the normal. The anterior pituitary, on the contrary, has a spurt in weight growth at adolescence in both sexes, and a sex difference appears with the female gland in this instance growing larger than the male and acquiring a higher percentage of acidophils (Rasmussen, 1947, 1950). In rats both increased

anode American Phillips. We installed a small dark-room, and began working at techniques for displaying bone muscle and fat in the legs and arms, with the expert eyes of Dr. T. Hills and Mr. R. W. Stanford of Guy's Hospital constantly upon us, the [illegible]

graphers, Mrs. Smith and Mrs. Cutler, who currently take radiographs of the upper arm, thigh, calf and chest of each child, while Mr. Parfitt takes the wrist and hand for skeletal maturation ratings and the jaws for tooth eruption and root maturation data. A few months ago Mr. Ballard, also of the Eastman Dental Clinic, became interested in the possibilities our children presented for orthodontic research, and we are now beginning to take lateral skull films as well. In addition to this work, and somewhat more peripheral to our team, Mr. W. H. Hammond, of the Ministry of Health, is collating school records and intelligence tests on the children, and Miss Washington, of the Ministry, is undertaking periodic nutritional surveys of the constituent houses that make up the Home.

Techniques in use. The main team, now numbering nine people, spends two days at the Home every month. Each child is seen every six months, and every three months during puberty, and by going every month we make sure of measuring the child within fifteen days of his birthday, half-birthday or quarter-birthday. In our results we correct for the not-seen-on-birthday error by linear interpolation We see about twenty children on each of the two days, and deal with either three or four children every hour.

The child undresses and is first measured. We measure stature, weight, sitting height (Martin anthropometer with the child sitting on a table as described in Hooton, 1946), biacromial diameter, bi-iliac diameter, bi-condylar diameters of humerus and femur, and the subcutaneous tissue folds over the biceps, over the triceps, under the left scapula and above the left iliac crest The child is then posed for the photogrammetric picture, which is taken with an F.24 aircraft camera fitted with a 20-in lens 10 m distant from the centre of the turntable on

Rev. J. M. Waterhouse, Principal of the National Children's Home, and Mr. E. Shutt, the Governor of the Harpenden Branch. As for the children, they *are* the Study, and if one can scarcely call some of them models of patience, they have all the other scientific virtues. They are for ever testing the range of our equipment, both physical and mental, with all the means at their command, and take a peculiar delight in detecting any slovenliness of technique on our part; and they possess in abundance those high spirits that go with the knowledge of a job well done, and an afternoon out of school to do it in.

There are some 250 children in the Home, some orphans, others from broken and disintegrated homes. Two-thirds are boys, one-third girls, and practically all participate in the Study. Their ages range from 3 to 18, but as the majority are between 5 and 16 we lack a very young group, and we tend to lose the older ones before their growth is quite completed; we are endeavouring to follow them up after they leave, but whether we shall be successful in this I do not know. The Study has been built up gradually over the last three years. The first to join our team, in November 1948, was Mr. R. H. Whitehouse, our anthropometrist and administrative secretary. Then, because in addition to the usual anthropometry we wished to have specialized serial photographs of the children that could also be measured, we enlisted the help of the Photographic and Reproduction Branch of the Air Ministry. They helped to solve our photographic problems, and for a considerable period undertook the arduous task of processing our pictures. In June 1949 we started operations in an old disused laundry, but a few months later the greatest stroke of good fortune befell us and we were able to move to a permanent set-up on quite a large scale in one of the buildings of the Home. In October we were joined by Dr. Cecile Asher from the Ministry of Education, who took over the clinical examinations and medical histories, and in September of the following year by Mr. G. Parfitt from the Eastman Dental Clinic who began a long-term study of tooth eruption, root maturation and caries development. Meanwhile we had acquired two war-surplus X-ray machines, one dental and the other a 200 mA stationary-

there to the main radiographic room. The technique used here is illustrated in Figs. 16 and 17. The anode is placed 2·5 m away from the film to minimize magnification and the uncorrectable error which occurs if the maximum diameter of the limb is not at exact right angles to the central beam. We would

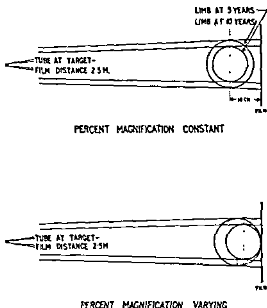


FIG. 16. Diagram of radiographic technique, to show (above) how keeping the central plane of the limb a constant distance from the casette gives a constant magnification factor, rather than (below) an increasing one if the limb is always in contact with the casette.

like to have an even larger distance, so as to achieve the accuracy possible in the photographs, but the limitations of our equipment prevent this The central plane of the limb is placed a fixed distance away from the film (10 cm for calf and thigh; 5 cm for arm) so that the magnification stays constant rather than increasing as the child grows larger, as it would if the limb was every time in contact with the casette (Fig. 16). The way

which the subject stands (Tanner and Weiner, 1949, appendix; Tanner, 1951b). The child's heels are placed 8 cm behind the centre of the turntable if he is less than 10 years old, and 10 cm if he is over 10, and the pose is that recommended by Dupertuis

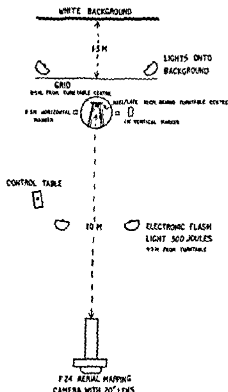


FIG. 14. Diagram of the apparatus and technique used for photogrammetry.

and Tanner (1950). The background is illuminated to prevent shadow, a grid is fixed behind the child to give horizontals and verticals for help in posing and in measuring the picture and to give a visual impression of size increase in serial photographs, and electronic flash lighting is used. The set-up is illustrated in Fig. 14, and a typical set of pictures is shown in Fig. 15.[1]

After this, the child has his wrist and jaw radiographs taken and his teeth examined, and then goes to the clinician, and from

[1] For Figs 15 and 17 see Plates V and VI following p 396.

wards animals and the possibility of animal experimentation. But here we run straight into, on the one hand, trouble; on the other, interest. It seems that the adolescent spurt, at least in the sense in which we have been discussing it, is the perquisite of primates. Other animals manage their reproductive affairs differently, and with considerably more dispatch. I am aware this is a very sweeping statement; exceptions may well exist, and the data are anyway too flimsy to support such a generalization with certitude. Nevertheless the evidence, some of which I will now show you, does seem to point towards this conclusion.

Practically all we have to go on are curves for body weight, [illegible] own warning [illegible] weight curves [illegible] anner, 1951a). However, there is nothing else for it until longitudinal data on linear dimensions of animals become available. At present growth data on animals are exceedingly scanty and where existent suffer even more acutely than the human material from the evils of cross-sectional reporting of longitudinal data (Tanner, 1951a). The paucity of records is something of a novelty to the human biologist, who is more accustomed to a variety of digestive ailments than to starvation. There are a few reports on physiological changes during growth, notably in muscle constituents (Horvath, 1945) and serum alkaline phosphatase (Li, Kalman, Evans and Simpson, 1946) in rats For linear measurements we have fairly extensive postnatal figures for cattle and pigs given in Brody, Hogan, Kempster, Ragsdale and Trowbridge (1926) and reprinted (for cattle only) in Brody (1945), and prenatal figures for cattle (Winters, Green and Comstock, 1942) and sheep (Winters and Feuffel, 1936; Barcroft, 1946). In rats the most extensive data are those of Outhouse and [illegible]

Saxton and Silbeberg (1947) and Barnes, Sperling and McCay (1947) on body and tibia lengths, and Stotsenberg (1915) on prenatal body length. For mice there are body, tibia and third

a 6-year-old rising to 55 in a 16-year-old, with a mAs of 12 for arm and calf and 20 for thigh. Though films taken direct on to Ilfex film are considerably superior to others in detail, we have to use par speed screens at Harpenden because of the increase of exposure needed for the Ilfex. Chests we take primarily for clinical purposes, and are very fortunate in having them read by Dr. J. W. Peirce of St. Thomas's Hospital. Mr. Stanford has measured our total dose of X-rays to the gonads of boys as 3·5 mr per attendance and to girls as 2 mr per attendance. About the maximum number of attendances we expect any child to have is 30, so that in all we might give boys a gonad dose of 105 mr and girls 60 mr. An estimate of the dose of X-rays received during the first thirty years of life for ordinary diagnostic purposes by the average member of the population is about 100 mr for men and 300 mr for women. We are accordingly fairly confident that we are causing no appreciable increase in mutations in our children, and, of course, are nowhere near the level of effects on any other tissue of the child.

For following the morphological growth of the children, then, we have the anthropometric measurements, the photogrammetric picture, the soft-tissue radiographs, the wrist and jaw radiographs and the clinical description. I have no hesitation in saying that so far as this side goes the Harpenden Growth Study bears comparison with any study anywhere. But this I regard as only a basis for development of the biochemical and physiological side, and here we are as yet severely limited. We hope soon to be able to add biochemical help to the team: meanwhile all we have so far done is to provide a little endocrinological material for Dr. F. T. G. Prunty and Dr. B. E. Clayton of St. Thomas's Hospital, in the way of serial gonadotrophin and 17-ketosteroid determinations at adolescence. And that brings me back to the main theme of this lecture, and to our final subject, the evolutionary position of the adolescent spurt.

THE ADOLESCENT SPURT IN ANIMALS

While we were discussing the endocrinology of adolescence there were so many speculations and so much data required to test them, that the minds of readers must have turned to-

in mind. In them are charted the velocity curves of weight from establishment of the foetus to physical maturity in guinea-pigs, mice, rats, sheep and cattle. I have endeavoured to choose the best data, where a choice was possible, bearing in mind particularly nutritional conditions and the relative longitudinality of the data. Some small indication that differences in nutrition

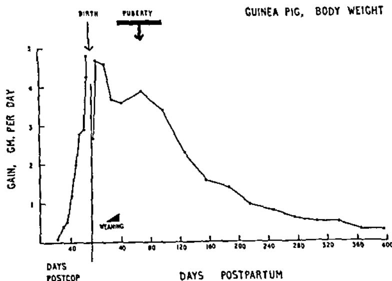


Fig. 18. Weight velocity curve for the guinea-pig. Prenatal data, sexes combined, cross-sectional, from Draper (1920); postnatal, pure longitudinal on 770 males, from Sewall Wright cited in Brody, Hogan, Kempster, Ragsdale and Trowbridge (1926). Time of puberty from Young, Dempsey, Hagquist and Boling (1939), first heat in females approx. 65 days, stand. dev. 21 days.

may not affect the general shape of these curves can be seen in Fig. 20, where curves 4, 7 and 8 are from 1941 data now known to be suboptimal, and curves 2, 3, 5, and 6 are from 1948 data. Though the magnitude of the values is different, the shape of the 1941 and 1948 curves is very much the same.

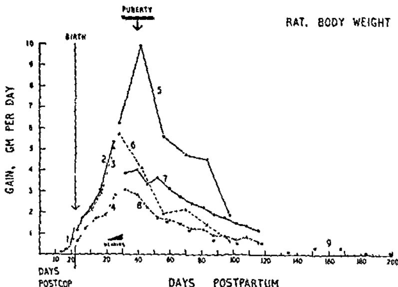
The main features of the weight velocity curves of the five species are similar. The curve rises to a peak and then falls more or less smoothly towards zero level. Birth occurs before the peak is reached, either a considerable time before as in rats and

metacarpal lengths for a small group of C57 blacks (Green and Fekete, 1933) and a few figures for body, ear, foot and tail length given by MacArthur and Chiasson (1945). Draper (1920) and Bell (1941) give lengths of prenatal guinea-pigs, and metacarpal and metatarsal lengths in rabbits over a short age range are given by Crary and Sawin (1949). All these data except the last are reported cross-sectionally.

For body weight the situation is considerably better, but the data are still by no means comparable in quality to those on the human. Growth studies in animals have their own peculiar difficulties: the technique of measuring animals is quite undeveloped, and faces greater difficulties than anthropometry, because animals possess fur and because they cannot co-operate like children during measurement. Differing nutritional conditions play a far greater part in animal growth than in human, where the children of the major studies are under nearly optimal conditions of nourishment. We are only now beginning to find out how to feed laboratory animals optimally; the laboratory mouse and rat have been growing larger and faster every decade and this trend still shows no sign of stopping. As for farm animals, they are deliberately undernourished for economic reasons, and only from the very few research herds can one expect data comparable to the human. Then there is the difficulty of litter size in the smaller animals, with weight and weight increments being smaller for many-membered litters, due chiefly it seems to the mothers' inability to supply enough food for so many months during pregnancy and certainly after it Weaning in these animals is also often associated with a loss of weight due to the mothers' milk supply decreasing before outside foods can be taken sufficiently This happens sometimes in the human as well, but in the laboratory animals weaning occurs at a time that is much more critical from the point of view of the investigator of growth processes, as I shall make clear in a moment. And lastly there is the varying establishment of pubescence in species which are seasonal breeders, and the great difficulty of obtaining pure longitudinal birth-to-maturity records due to the high mortality of most laboratory animals.

Figs. 18 to 22 must be looked at with all these difficulties held

of which only four are plotted. This may represent a miniature adolescent spurt: it is certainly not a weaning phenomenon. There is little evidence of a similar thing in the rat data, unless the kink in curve 7 be taken as such. The sheep and cattle are



insufficiently precise to show anything on so small a scale; the rise in the sheep curves after puberty is probably due to pasturing changes. It may be that in these laboratory and domestic animals some small increase of velocity occurs about the time of puberty—indeed since the gonads grow rapidly at this time there must be a tiny increase unless perfectly balanced by fat loss—but there is no large adolescent spurt as in the human. Some of the larger wild mammals may possibly approach the

cattle or not very long before, as in guinea-pigs and sheep. Puberty, judged in the rodents by first oestrus, occurs after the peak is passed, but as a rule not very long after. There does not seem to be any particular rise in weight velocity associated with puberty, but we should be very cautious about this statement.

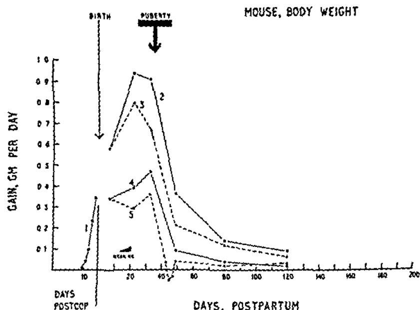


Fig 19 Weight velocity curve for the mouse. Curve 1 sexes combined,

l by
and
tler
27)

The rise seen at about this time in the guinea-pig data, and in some of the rat and mouse curves, is more likely to be the effect of weaning difficulties, immediately before weaning the

ever, *Butler and Metrakos (1950) indicate that the transient dip* followed by the rise at 42–50 days seen plotted in Fig. 19, curve 5, occurred in most or all their eighteen groups of mice,

in children starved during the midgrowth years and fed more properly about the time of puberty (Brody, 1945, pp. 511, 539). This is, of course, a totally erroneous notion, bred of the fact that animals when starved slow down in growth, and when re-fed accelerate back to their pre-starvation, constitutional growth curve (Outhouse and Mendel, 1933; Barnes, Sperling and

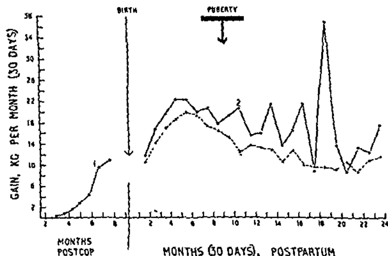


Fig. 22. Weight velocity curve for cattle Curve 1, sexes combined, cross-sectional, from Winters, Green and Comstock (1942). Curves 2 and 3, males and females, mixed longitudinal, Jersey cattle, from Brody *et al.* (1926). Puberty approximate, from Asdell (1946).

McCay, 1947), as indeed do humans, probably at any time of their growth (Wolff, 1940).

The postponement and magnification of adolescence, however, are not perquisites of the human alone; in Fig. 24 are shown the curves for the chimpanzee (Grether and Yerkes, 1940), and they are in all respects closely similar to man's, even down to the sexual differences. The female spurt is less marked than the male, and possibly a trifle earlier, too, as in the human The adolescent spurt proper is apparently an evolutionary step taken by the primates. Unfortunately there are no longitudinal growth data available on monkeys, but it is known that in

human more closely. The crucial data, correlating weight growth and linear growth with first oestrus or with testis growth in individual animals, are completely lacking for any species.

When the curves for these laboratory and domestic animals are compared with that of the human in Fig. 23, a vast difference

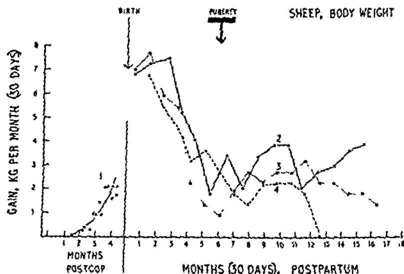


Fig 21. Weig[illegible]
sectional, from [illegible]
Curves 2, 3 a[illegible]
Merino females [illegible]
dinal, from Fos[illegible]
Asdell (1946).

ence is at once seen. In the human curve the initial peak is there, with birth occurring shortly before or coincident with it, and weaning would be in about the same relative position as in the other species. But the time at which puberty occurs is enormously retarded, and when it does occur, it is associated with a marked increase in weight velocity which produces a distinct second peak in the curve. This is such a radically different picture from the laboratory and domestic animals that a few authors, lacking experience of the human, have rashly jumped to the conclusion that the adolescent spurt is an artefact occurring only

this difference represents the small pre-adolescent difference to be seen in primates. The anabolic effect of androgens is not confined to the primate, and there would certainly seem to be differences of shape in many species as well as in such things as the lion's mane and the stag's antlers. We know too little about the morphology and physiology of mammals in their normal state,

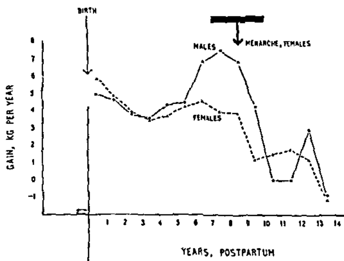


FIG 24. Weight velocity curve for the chimpanzee. Actual increments, from Grether and Yerkes (1940).

however, to characterize the morphology more closely, or to say how far the functional differences which occur between men and women are present amongst other mammals also.

The cause of the postponement of puberty lies, it seems, in the primate hypothalamus. The increased time necessary for the maturing of the primate brain has been sandwiched in between weaning and puberty; the maturation of the hypothalamus has been put back so that maturation of the association areas of the cortex and other parts can take place first. Presumably the biological reason for this is that it is relatively useless to produce young while the brain is still immature, as the advantage the brain gives in their protection will not then occur. The literature

Macaca mulatta the weight of males and females is the same before puberty, while the male is 20 per cent heavier in mature animals (Kennard and Willner, 1941).

In this connexion all sorts of questions immediately present themselves. Is the spurt merely seen because postponement of puberty makes it obvious? Puberty in other mammals occurs

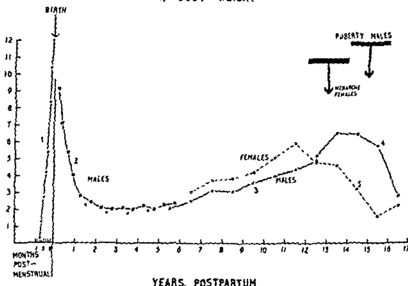


Fig. 23 Weight velocity curve for man. Curve 1, sexes combined, cross- [illegible]

from Simmonds (1944).

shortly after the maximum velocity of weight growth has been passed, and often when growth is still proceeding fairly vigorously; a small spurt then might be swamped in the general growth. Other mammals also tend to go on growing for a good deal longer after puberty, relatively speaking, than the primate, giving the impression that the essential change is a postpone- [illegible]

[illegible] greater in [illegible] than the female in most mammals, but is so at birth, and it is possible that

REFERENCES

BACKMAN, G. (1948). Die beschleunigte Entwicklung der Jugend. Verfruhte Menarche, verspätete Menopause, verlängerte Lebensdauer *Acta anat.* 4, 421.

BARCROFT, J. (1946). *Researches on pre-natal life*, vol. i. Blackwell, Oxford.

BARNES, L. L., SPERLING, G., and MCCAY, C. M. (1947). Bone growth in normal and retarded growth rats *J. Gerontol* 2, 240.

phys Anthrop, N S. 4, 433.

BECKS, H., ASLING, C. W., COLLINS, D A., SIMPSON, M. E, and EVANS, H. M.

BLISS, C. I., and YOUNG, M S. (1950) An analysis of heart measurements of growing boys *Hum. Biol.* 22, 271.

BOAS, F (1892). The growth of children. *Science*, 19, 256; 19, 281; 20, 351.

— (1932) Studies in growth. *Hum Biol* 4, 307.

— (1933) Studies in growth, II. *Hum Biol* 5, 429.

— (1935). Studies in Growth, III. *Hum Biol* 7, 303

BOYNTON, B (1936) The physical growth of girls A study of the rhythm of physical growth from anthropometric measurements on girls between birth and 18 years *Univ Iowa Stud. Child Welf* 12, No. 4.

BRODY, S (1945). *Bioenergetics and Growth.* Reinhold, New York.

— HOGAN, A. G, KEMPSTER, H L, RAGSDALE, A. C., and TROWBRIDGE, E A. (1926) Growth and development, with special reference to domestic animals I Quantitative data 1 Weight growth and linear growth *Univ. Missouri Agricult exp Sta. Res Bull.* No 96

*on precocious puberty makes it clear that the adolescent mech*anism, once started, is readily capable of going all the way to spermatogenesis and corpora lutea, even at the age of 5 or 6. The hypothalamus alone, indeed, stands between us and six-year-old pregnancy. The postponement of adolescence is a highly distinctive evolutionary trait in primates, and a necessary one, if the long maturation period for the brain is to be taken advantage of.

Such are some of the evolutionary problems raised by the adolescent spurt. The psychological and sociological problems I have not even touched on. The genetic background I have barely mentioned. The unsolved problems clustering around human growth range all the way from the most fundamental [illegible] immedi- [illegible] seems to me that just as sixty years ago Boas realized that further advances then depended on the setting up of longitudinal growth studies, so now we should realize that growth studies of man must be merged into the wider study of human biology as a whole. What we need—and we need it as the basis for a humanistic medicine, as much as for anything else—is a concerted attack in this field by anthropologists, physiologists, zoologists, physicians. We need to have our future Growth Study in an Institute of Human Biology, wherein study of the human does not cease when he is sixteen or seventeen, but follows him and his children and his grandchildren, in the same spirit and with the same intention as the great humanistic physicians brought to their calling.

ACKNOWLEDGEMENTS

The figures were drawn by Mr R. H. Whitehouse, to whom my thanks are due for this and much other indispensable assistance. I wish also to thank Miss Lauretta Stokes for her care and persistence with the typescript

FELDMAN, J. D (1951). Endocrine control of lymphoid tissue *Anat Rec.* **110**, 17.
FITT, A. B (1941). *Seasonal influence on growth function and inheritance* Wellington, N Z.: New Zealand Council for Educational Research
FLORY, C D. (1936). Osseous development in the hand as an index of
C
fifteen years of observation. *Arch Neurol. Psychiat., Chicago*, **41**, 755.
GOULD, H. N, and GOULD, M. R. (1932). Age of first menstruation in mothers and daughters. *J. Amer. med. Ass.* **98**, 1349.
GREEN, C V, and FEKETE, E. (1933). Differential growth in the mouse. *J. exp. Zool* **66**, 351.
GRETHER, W F., and YERKES, R. M. (1940). Weight norms and relations for chimpanzee. *Amer. J. phys Anthrop.* **28**, 181.
—
—
—
individual girls before, during, and after puberty. *Yale J Biol. Med.* **17**, 91.
HAMBURGER, C. (1948) Normal urinary excretion of neutral 17-ketosteroids. With special reference to age and sex variation. *Acta endocr. Copenhagen*, **1**, 19.
— HALVORSEN, K, and PEDERSEN, J. (1945) Assay of androgenic substances in the urine of normal men and women *Acta pharm. tox. Kbh* **1**, 129.
HAMILTON, J. B. (1951). Quantitative measurement of a secondary sex character, axillary hair. *Ann. N Y. Acad. Sci.* **53**, 585
HARRIS, G W., and JACOBSOHN, D (1952). Functional grafts of the anterior
I
I
[illegible] M. (1945). The distribution of phosphorous compounds in the gastrocnemius muscle as influenced by the ageing process *Amer. J. Physiol.* **145**, 77
JOHNSTON, J A (1941). Factors influencing retention of nitrogen and calcium in period of growth. IV. Effect of estrogen *Amer. J Dis Child* **62**, 708.
JONES, H. E. (1949). *Motor performance and growth.* University of California
[illegible] *Amer J. Roentgenol.* **57**, 477.

BRUEN, C. (1933). Variation of basal metabolic rate per unit surface area II. Pubertal acceleration *J. Nutr.* 6, 383

BUTLER, L., and METRAKOS, J. D (1950). A study of size inheritance in the house mouse. I. The effect of milk source. *Canad. J. Res. Sec. D.* 28, 16

CASTALDI, L , and VANUCCI, D (1927). Le misure antropometriche esterne e i pesi viscerali piu importanti considerati in funzione del sesso, età, statura e costituzione. *Scr. biol. Castaldi.* 1, 151.

CHRISTIE, A. (1949). Prevalence and distribution of ossification centres in the newborn infant. *Amer. J. Dis. Child.* 77, 355

CLARK, L. C., and BECK, E. I. (1950). Plasma alkaline phosphatase activity I. Normative data for growing children *J. Pediat.* 36, 335.

— THOMPSON, H. L., BECK, E. I , and JACOBSON, W. (1951). Excretion of creatine and creatinine by children. *Amer. J. Dis. Child.* 81, 774.

COLE, H. H , and CASADY, R. B. (1947). Studies on indices and causes of

Genetics, 34, 508.

DAVENPORT, C. B. (1934). Critique of curves of growth and of relative growth. *Cold Spr. Harb. Symp quant. Biol.* 2, 203.

DEAMER, W C. (1948). Stimulation of growth in boys by sublingual testosterone therapy. *Amer J. Dis. Child.* 75, 850.

DIMOCK, H. S (1935). Research in adolescence I. Pubescence and physical growth. *Child Develpm* 6, 177.

DOWNING, M (1947) Blood pressure on normal girls from 3 to 16 years of age. *Amer. J. Dis. Child* 73, 293

DRAPER, G , DUPERTUIS, C .W., and CAUGHEY, J. L. (1944). *Human constitution in clinical medicine.* Hoeber, New York.

DRAPER, R. L. (1920). The prenatal growth of the guinea-pig *Anat. Rec.* 18, 369

DUPERTUIS, C. W., and TANNER, J. M. (1950). The pose of the subject for photogrammetric anthropometry, with special reference to somatotyping. *Amer. J phys Anthrop.*, N S. 8, 27.

EMERSON, G. A., and EVANS, H M (1944) The bioassay of vitamin E. *J. Nutr.* 27, 469

ENGLE, E. T , and ROSASCO, J. (1927). The age of the albino mouse at normal sexual maturity *Anat Rec.* 36, 383.

ENGSTROM, W W , and MUNSON, P. L. (1951). Precocious sexual and somatic development in boys due to constitutional and endocrine factors. *Amer. J. Dis. Child.* 81, 179.

ESPENSCHADE, A (1950). Motor performance in adolescence. *Monogr. Soc. Res Child Develpm* 5, No 1

FELDMAN, J. D. (1951). Endocrine control of lymphoid tissue. *Anat Rec.* 110, 17.

FITT, A. B (1941). *Seasonal influence on growth function and inheritance.* Wellington, N.Z.: New Zealand Council for Educational Research.

FLORY, C. D. (1936). Osseous development in the hand as an index of skeletal development. *Monogr. Soc. Res. Child Develpm.* 1, No. 3.

GESELL, A., THOMS, H., HARTMAN, F. B., and THOMPSON, H. (1939). Mental [illegible]

G[illegible] mothers and daughters. *J. Amer. med. Ass* 98, 1349.

GREEN, C. V., and FEKETE, E. (1933). Differential growth in the mouse. *J. exp. Zool.* 66, 351.

GRETHER, W. F., and YERKES, R. M. (1940) Weight norms and relations for chimpanzee. *Amer. J. phys. Anthrop.* 28, 181.

GREULICH, W. W., DORFMAN, R. I., CATCHPOLE, H R., SOLOMON, C. I , and [illegible]

HAMBURGER, C. (1948). Normal urinary excretion of neutral 17-ketosteroids. With special reference to age and sex variation *Acta endocr. Copenhagen,* 1, 19.

— HALVORSEN, K., and PEDERSEN, J (1945). Assay of androgenic substances in the urine of normal men and women *Acta pharm tox Kbh* 1, 129

HAMILTON, J B. (1951). Quantitative measurement of a secondary sex character, axillary hair. *Ann. N.Y. Acad. Sci.* 53, 585.

HARRIS, G. W., and JACOBSOHN, D. (1952). Functional grafts of the anterior [illegible]

HOOTON, E. A. (1946). *Up from the Ape.* Revised ed. Macmillan, New York.

HORVATH, S M. (1945). The distribution of phosphorous compounds in the gastrocnemius muscle as influenced by the ageing process *Amer. J. Physiol.* 145, 77.

JOHNSTON, J. A. (1941). Factors influencing retention of nitrogen and calcium in period of growth. IV. Effect of estrogen. *Amer J Dis. Child.* 62, 708.

[illegible] infants. *Amer. J. Roentgenol.* 57, 477.

KENNARD, M. A., and WILLNER, M. D. (1941). *Findings in 216 routine autopsies of Macaca mulatta. Endocrinology*, 28, 955.

KINSELL, L. W., MICHAELS, G D , LI, C. H , and LARSEN, W. E. (1948). Studies in growth. I. Interrelationship between pituitary growth factor and growth-promoting androgens in acromegaly and gigantism II. Quantitative evaluation of bone and soft tissue growth in acromegaly and gigantism *J. clin. Endocr.* 8, 1013.

KORENCHEVSKY, V., DENNISON, M., and SIMPSON, S. L. (1935). The prolonged treatment of male and female rats with androsterone and its derivatives, alone or together with oestrone. *Biochem. J* 29, 2534.

KROGMAN, W. M (1941a). Growth of man *Tabul. biol. Amst.* 20, 1.

— (1941b). *Bibliography of Human Morphology 1914–1939.* Chicago University Press.

— (1950). The physical growth of the child. syllabus. *Yearbook of phys. Anthrop.* 1949. 5, 280.

LE MARQUAND, H. S., and RUSSELL, D. S. (1934–35). A case of pubertas praecox (macrogenitosomia praecox) in a boy associated with a tumour in the floor of the third ventricle *Royal Berkshire Hosp. Rep.* 3, 11.

LEWIS, R. C , DUVAL, A. M , and ILIFF, A. (1943a). Basal metabolism of normal children from 13 to 15 years old, inclusive. *Amer. J. Dis. Child* 65, 845.

— DUVAL, A M , and ILIFF, A (1943b). Effect of adolescence on basal metabolism of normal children *Amer. J. Dis. Child* 66, 396.

LI, C. H , KALMAN, C., EVANS, H M., and SIMPSON, M. E. (1946). The effect of hypophysectomy and adrenocorticotropic hormone on the alkaline phosphatase of rat plasma *J. biol Chem.* 163, 715

LISSER, H , CURTIS, L E , ESCAMILLA, R F., and GOLDBERG, M. B. (1947). The syndrome of congenitally aplastic ovaries with sexual infantilism, high urinary gonadotropins, short stature and other congenital anomalies *J. clin Endocr* 7, 665

MCARTHUR, J W , and CHIASSON, L. P (1945). Relative growth rates in races of mice produced by selection. *Growth*, 9, 303

MACDOWELL, E. C., ALLEN, E , and MACDOWELL, C G (1927). The prenatal growth of the mouse *J gen Physiol.* 11, 57.

MARESH, M. (1948) Growth of the heart related to bodily growth during childhood and adolescence *Pediatrics*, 2, 382

MAYER, J (1948). Growth characteristics of rats fed a synthetic diet *Growth*, 12, 341

MENEES, T. O , and HOLLY, L. E (1932) Ossification in extremities of new born *Amer J Roentgenol* 28, 389

MEREDITH, H. V. (1935) The rhythm of physical growth a study of 18 anthropometric measurements on Iowa City white males ranging in age between birth and 18 years *Univ Iowa Stud Child Welf* 11, No 3

— (1946). Order and age of eruption for the deciduous dentition *J dent. Res.* 25, 43

MILLS, C. A , and OGLE, C. (1936). Physiological sterility of adolescence. *Hum Biol* 8, 607.

MOMENT, G. B (1933). The effects of rate of growth on the post-natal [illegible]

[illegible]

Growth, 12, 311.

NATHANSON, I. T., TOWNE, L. E , and AUB, J. C. (1941). Normal excretion of sex hormones in childhood. *Endocrinology*, 28, 851

NYDA, M. J., DE MAJO, S. F., and LEWIS, R. A. (1948) The effect of ovariectomy and physiologic doses of estradiol upon body weight, linear growth and fat content of the female albino rat. *Johns Hopk. Hosp. Bull.* 83, 279

NYLIN, G. (1935). The physiology of the circulation during puberty. *Acta. med. scand.*, Suppl 69

OESTING, R. B , and WEBSTER, B (1938). The sex hormone excretion of children. *Endocrinology*, 22, 307.

OUTHOUSE, J., and MENDEL, L. B (1933). The rate of growth. I. Its influence on the skeletal development of the albino rat. *J. exp Zool.* 64, 257.

PALMER, C. E. (1933). Seasonal variation in growth of elementary school children *Pbl. Hlth. Rep. Wash.* 48, 211.

— KAWAKAMI, R , and REED, L. J. (1937) Anthropometric studies of individual growth. II. Age, weight and rate of growth in weight, elementary school children. *Child Develpm.* 8, 47.

— and REED, L. J. (1935) Anthropometric studies of individual growth I. Age, height and growth in height, elementary school children. *Hum. Biol* 7, 319

PEDERSEN-BJERGAARD, K , and TØNNESEN, M. (1948a). Oestrogenic, androgenic and gonadotrophic substances in the urine of normal women Sex hormone analyses *Acta endocr Copenhagen*, 1, 38.

— and TØNNESEN, M (1948b). Sex hormone analyses II. The excretion of sexual hormones by normal males, impotent males, polyarthritics and prostatics *Acta. med scand.*, Suppl 213, 284.

PETRI, E (1935) Untersuchungen zur Erbbedingtheit der Menarche. *Z Morph Anthr.* 33, 43

POPENOE, P (1928) Inheritance of age of onset of menstruation *Eugen News*, 13, 101.

PREISEL, R., and WAGNER, R (1931) Gesetzmassigkeiten im Auftreten der extragenitalen sekundaren Geschlechtsmerkmale bei Madchen. *Z ges. Anat* , 2. *Z KonstLehre* 15, 333

PRYOR, H. B. (1936). Certain physical and physiological aspects of adolescent development in girls. *J Pediat.* 8, 52

RASMUSSEN, A T (1947) The growth of the hypophysis cerebri (pituitary gland) and its major subdivisions during childhood *Amer. J. Anat* 80, 95.

KENNARD, M. A., and WILLNER, M. D. (1941). Findings in 216 routine autopsies of *Macaca mulatta*. *Endocrinology*, **28**, 955.

KINSELL, L W., MICHAELS, G. D., LI, C. H., and LARSEN, W. E. (1948). Studies in growth. I. Interrelationship between pituitary growth factor and growth-promoting androgens in acromegaly and gigantism. II. Quantitative evaluation of bone and soft tissue growth in acromegaly and gigantism *J. clin. Endocr.* **8**, 1013.

KORENCHEVSKY, V., DENNISON, M., and SIMPSON, S L. (1935). The prolonged treatment of male and female rats with androsterone and its derivatives, alone or together with oestrone. *Biochem. J* **29**, 2534.

KROGMAN, W. M. (1941a). Growth of man. *Tabul. biol. Amst* **20**, 1.

— (1941b). *Bibliography of Human Morphology 1914–1939* Chicago University Press.

— (1950). The physical growth of the child: syllabus. *Yearbook of phys. Anthrop. 1949*. **5**, 280

LE MARQUAND, H S., and RUSSELL, D. S. (1934–35). A case of pubertas praecox (macrogenitosomia praecox) in a boy associated with a tumour in the floor of the third ventricle *Royal Berkshire Hosp. Rep.* **3**, 11.

LEWIS, R. C, DUVAL, A. M., and ILIFF, A. (1943a). Basal metabolism of normal children from 13 to 15 years old, inclusive. *Amer. J. Dis. Child.* **65**, 845.

— DUVAL, A. M., and ILIFF, A (1943b). Effect of adolescence on basal metabolism of normal children. *Amer. J. Dis. Child.* **66**, 396.

LI, C. H., KALMAN, C., EVANS, H. M., and SIMPSON, M. E (1946). The effect of hypophysectomy and adrenocorticotropic hormone on the alkaline phosphatase of rat plasma *J. biol Chem* **163**, 715.

LISSER, H., CURTIS, L. E., ESCAMILLA, R. F., and GOLDBERG, M B. (1947). The syndrome of congenitally aplastic ovaries with sexual infantilism, high urinary gonadotropins, short stature and other congenital anomalies. *J. clin Endocr* **7**, 665.

MCARTHUR, J. W., and CHIASSON, L P. (1945). Relative growth rates in races of mice produced by selection *Growth*, **9**, 303

MACDOWELL, E. C., ALLEN, E , and MACDOWELL, C. G. (1927). The prenatal growth of the mouse *J gen Physiol* **11**, 57.

MARESH, M. (1948) Growth of the heart related to bodily growth during childhood and adolescence. *Pediatrics*, **2**, 382.

MAYER, J. (1948). Growth characteristics of rats fed a synthetic diet. *Growth*, **12**, 341

MENEES, T O., and HOLLY, L E. (1932) Ossification in extremities of new born. *Amer J. Roentgenol* **28**, 389

MEREDITH, H. V (1935). The rhythm of physical growth a study of 18 anthropometric measurements on Iowa City white males ranging in age between birth and 18 years *Univ Iowa Stud Child Welf* **11**, No. 3.

— (1946). Order and age of eruption for the deciduous dentition *J. dent. Res.* **25**, 43

MILLS, C. A, and OGLE, C. (1936). Physiological sterility of adolescence. *Hum. Biol.* 8, 607.

MOMENT, G. B. (1933). The effects of rate of growth on the post-natal

Growth, 12, 311.

NATHANSON, I T., TOWNE, L. E., and AUB, J. C. (1941). Normal excretion

The effect of ovari-
body weight, linear growth and fat content of the female albino rat. *Johns Hopk. Hosp. Bull.* 83, 279.

NYLIN, G. (1935). The physiology of the circulation during puberty. *Acta. med scand*, Suppl. 69

OESTING, R. B, and WEBSTER, B (1938) The sex hormone excretion of children *Endocrinology*, 22, 307.

OUTHOUSE, J, and MENDEL, L. B. (1933) The rate of growth. I. Its influence on the skeletal development of the albino rat. *J exp. Zool* 64, 257

PALMER, C E. (1933). Seasonal variation in growth of elementary school children. *Pbl. Hlth Rep. Wash.* 48, 211

— KAWAKAMI, R., and REED, L. J (1937) Anthropometric studies of indi-

Biol. 7, 319.

PEDERSEN-BJERGAARD, K., and TONNESEN, M. (1948a). Oestrogenic, androgenic and gonadotrophic substances in the urine of normal women. Sex hormone analyses. *Acta endocr. Copenhagen*, 1, 38

— and TONNESEN, M. (1948b) Sex hormone analyses II. The excretion of sexual hormones by normal males, impotent males, polyarthritics and prostatics *Acta. med scand.*, Suppl 213, 284.

PETRI, E. (1935). Untersuchungen zur Erbbedingtheit der Menarche. *Z. Morph Anthr* 33, 43.

POPENOE, P. (1928). Inheritance of age of onset of menstruation *Eugen News*, 13, 101.

PREISEL, R, and WAGNER, R. (1931). Gesetzmassigkeiten im Auftreten der extragenitalen sekundaren Geschlechtsmerkmale bei Madchen *Z ges Anat.*, 2 *Z KonstLehre* 15, 333

PRYOR, H. B. (1936) Certain physical and physiological aspects of adolescent development in girls. *J Pediat.* 8, 52

RASMUSSEN, A T (1947) The growth of the hypophysis cerebri (pituitary gland) and its major subdivisions during childhood. *Amer. J Anat* 80, 95

RASMUSSEN, A. T. (1950). Changes in the proportion of cell types in the anterior lobe of the human hypophysis during the first 19 years of life. *Amer. J. Anat.* 86, 75.

REINHARDT, W. O., and WAINMAN, P. (1942). Effect of thyroidectomy, castration and replacement therapy on thymus, lymph nodes, spleen in male rats. *Proc. Soc. exp. Biol., N.Y.* 49, 257.

REYNOLDS, E. L. (1943). Degree of kinship and pattern of ossification. *Amer. J. phys. Anthrop.*, N.S. 1, 405.

— (1944). Differential tissue growth in the leg during childhood. *Child Develpm.* 15, 181.

— (1945). The bony pelvic girdle in early infancy. *Amer. J. phys. Anthrop.*, N.S. 3, 321.

— (1946). Sexual maturation and the growth of fat, muscle and bone in girls. *Child Develpm.* 17, 121.

— (1947). The bony pelvis in prepuberal childhood. *Amer. J. phys. Anthrop.*, N.S. 5, 165.

— (1949). The fat-bone index as a sex-differentiating character in man. *Hum. Biol.* 21, 199.

— (1951). The appearance of adult patterns of body hair in man. *Ann. N.Y. Acad. Sci.* 53, 576.

— and CLARK, L. C. (1947). Creatinine excretion, growth progress and body structure in normal children. *Child Develpm.* 18, 155.

— and GROTE, P. (1948). Sex differences in the distribution of tissue components in the human leg from birth to maturity. *Anat. Rec.* 102, 45.

— and Schoen, G. (1947). Growth patterns of identical triplets from 8 through 18 years. *Child Develpm.* 18, 130.

— and WINES, J. V. (1948). Individual differences in physical changes associated with adolescence in girls. *Amer. J. Dis. Child.* 75, 329.

— and WINES, J. V. (1951). Physical changes associated with adolescence in boys. *Amer. J. Dis. Child.* 82, 529.

RICHEY, H. G. (1931). The blood pressure in boys and girls before and after puberty. Its relation to growth and maturity. *Amer. J. Dis. Child.* 42, 1281.

ROBINOW, M. (1942). The variability of weight and height increments from birth to six years. *Child Develpm.* 13, 159.

RUSH, H. P., BILDERBACK, J. B., SLOCUM, D., and ROGERS, A. (1937). Pubertas praecox (macrogenitosomia). *Endocrinology*, 21, 404.

SANFORD, R. N., ATKINS, M. M., MILLER, R. B., and COBB, E. A. (1943). Physique, personality and scholarship. A co-operative study of school-children. *Monogr. Soc. Res. Child Develpm.* 8, No. 1.

SAXTON, J. A., and SILBERBERG, M. (1947). Skeletal growth and ageing in rats receiving complete or restricted diets. *Amer. J. Anat.* 81, 445.

SCAMMON, R. E. (1927). The first seriation study of human growth. *Amer. J. phys. Anthrop.* 10, 329.

— (1930). The measurement of the body in childhood. *In* HARRIS, J. A.,

JACKSON, C. M., PATERSON, D. G., and SCAMMON, R E., *The Measurement of Man* University of Minnesota Press.

SCHONFELD, W. A. (1943) Primary and secondary sexual characteristics Study of their development in males from birth through maturity, with biometric study of penis and testes. *Amer. J. Dis. Child.* **65**, 535

Dis. Child **78**, 484

SHOCK, N W. (1941). Age changes and sex differences in alveolar CO_2 tension. *Amer. J. Physiol* **133**, 610

— (1942) Standard values for basal oxygen consumption in adolescents. *Amer J Dis. Child.* **64**, 19

— (1943). The effect of menarche on basal physiological function in girls. *Amer. J. Physiol.* **139**, 288.

— (1944). Basal blood pressure and pulse rate in adolescents *Amer. J. Dis. Child.* **68**, 16.

— (1945) Creatine excretion in adolescents *Child Develpm.* **16**, 167.

— (1946a). Some physiological aspects of adolescence. *Texas Rep. Biol. Med* **4**, 289.

— (1946b). Physiological responses of adolescents to exercise. *Texas Rep. Biol Med.* **4**, 368

SHUTTLEWORTH, F. K (1934). Standards of development in terms of increments. *Child Develpm.* **5**, 89.

—

— (1939). The physical and mental growth of girls and boys age six to nineteen in relation to age at maximum growth *Monogr Soc. Res Child Develpm.* **4**, No. 3

— (1949a). The adolescent period a graphic analysis *Monogr. Soc. Res Child Develpm* **14**, No. 1.

— (1949b) The adolescent period: a pictorial atlas *Monogr. Soc. Res Child Develpm* **14**, No 2

SIMMONS, K (1944). The Brush Foundation study of child growth and development II. Physical growth and development. *Monogr Soc. Res Child Develpm* **9**, No 1

SIMMONS, K., and TODD, T. W. (1938). Growth of well children: analysis of stature and weight, 3 months to 13 years. *Growth*, 2, 93

STEARNS, G. A., and WARWEG, E. (1933). Studies of phosphorus of blood. I. The partition of phosphorus in whole blood and serum, the serum calcium, and plasma phosphatase from birth to maturity. *J. biol Chem.* 102, 749.

STOLZ, H. R., and STOLZ, L. M. (1951). *Somatic development of adolescent boys.* Macmillan, New York.

STONE, C. P., and BARKER, R. G. (1937). On the relationship between menarcheal age and certain measurements of physique in girls of ages 9 to 16 years. *Hum Biol.* 9, 1.

STOTSENBERG, J. M. (1915). The growth of the fetus of the albino rat from the thirteenth to the twenty-second day of gestation. *Anat. Rec.* 9, 667.

STREETER, G. L. (1920). Weight, sitting height, head size, foot length, and menstrual age of the human embryo. *Contr. Embryol Carneg. Instn.* 11, 143.

STUART, H C (1946). Normal growth and development during adolescence. *New England J. Med.* 234, 666; 693; 732

— and DWINELL, P. H. (1942). The growth of bone muscle and overlying tissues in children 6 to 10 years of age as revealed by studies of roentgenograms of the leg area. *Child Develpm.* 13, 195

— and REED, R. B. (1951). Certain technical aspects of longitudinal studies of child health and development. *Amer. J publ. Hlth* 41, 85

— and SOBEL, E. H. (1946) The thickness of the skin and subcutaneous tissue by age and sex in childhood. *J. Pediat* 28, 637.

TALBOT, F. B , WILSON, E B , and WORCESTER, J. (1937). Basal metabolism of girls, physiologic background and application of standards. *Amer. J. Dis. Child* 53, 275

TALBOT, N B , WOOD, M. S., WORCESTER, J., CHRISTO, E., CAMPBELL, A M., and ZYGMUNTOWICZ, A S (1951) Further observations on the urinary excretion of water-soluble corticosteroids by normal and abnormal subjects. *J clin Endocr.* 11, 1224

TANNER, J M. (1947). The morphological level of personality *Proc. roy Soc. Med.* 40, 301 Reprinted with additional figs. in *Yearbook phys Anthrop. 1947*, 3, 230

— (1948) A guide to American growth studies *Yearbook phys. Anthrop 1947*, 3, 28

— (1951a). Some notes on the reporting of growth data *Hum Biol* 23, 92

— (1951b). Current advances in the study of physique. Photogrammetric anthropometry and an androgyny scale *Lancet*, 1, 574

— (1951c). The relation between serum cholesterol and physique in healthy young men *J Physiol.* 115, 371.

— (1953) Inheritance of morphological and physiological traits In Sorsby, A., ed , *Clinical Genetics* Butterworth, London

— and WEINER, J S (1949) The reliability of the photogrammetric method of anthropometry, with a description of a miniature camera technique *Amer J. phys Anthrop* N S 7, 145

THOMPSON, D'A. W. (1942). *On Growth and form*. New ed. Cambridge University Press.

TOPPER, A., and MULIER, H. (1932). Basal metabolism of normal children· the puberty reaction. *Amer. J. Dis. Child.* 43, 327

TROLAND, C. E., and BROWN, C. A. (1948). Precocious puberty of intracranial origin. *J. Neurosurg* 5, 541.

VENNING, E H., and KAZMIN, V. (1946). Excretion of urinary corticoids and 17-ketosteroids in the normal individual *Endocrinology*, 39, 131.

VERSCHUER, O. v. (1931). Ergebnisse der Zwillingsforschung. *Anthrop Anz.* 6, 1.

WASHBURN, S. L. (1948). Sex differences in the pubic bone. *Amer J. phys. Anthrop.*, N.S 6, 199.

WEINBERGER, L. M, and GRANT, F. C. (1941). Precocious puberty and tumours of the hypothalamus. Report of a case and review of the literature, with a pathophysiologic explanation of the precocious sexual syndrome. *Arch. intern. Med* 67, 762.

WILKINS, L (1950). *The diagnosis and treatment of endocrine disorders in childhood and adolescence*. Thomas, Springfield.

WINTERS, L M, and FEUFFEL, G. (1936). Studies on the physiology of reproduction in the sheep IV. Fetal development *Univ Minnesota Agric. Exp Sta. Tech. Bull.*, No. 118.

— GREEN, W. W., and COMSTOCK, R. E (1942). Prenatal development of the bovine. *Univ. Minnesota Agricult. Exp. Sta Tech Bull*, No. 151

WOLFE, J M, WILSON, J. G, and HAMILTON, J B (1944). The effect of early postnatal injection of testosterone propionate on the structure of the anterior hypophysis of male and female rats *Yale J. Biol. Med.* 17, 341.

WOLFF, G. (1940). A study on the trend of weight in white school children from 1933–36 Material based on the examinations of pupils of the elementary schools in Hagerstown, Md. *Child Develpm* 11, 159

WOOD, M. E., and GRAY, C. H (1949). The urinary excretion of neutral 17-ketosteroids in childhood *J. Endocr.* 6, 111.

YOUMANS, J. B. (1948). Nutrition. I. Growth and development *J oral Surg, oral Med, and oral Pathol* 1, 168.

YOUNG, W. C, DEMPSEY, E W, HAGQUIST, C W, and BOLING, J L (1939) The sexual behaviour and sexual receptivity in the female guinea-pig *J. Comp Psychol.* 27, 49.

ZUCKER, L, HALL, L, YOUNG, M., and ZUCKER, T F (1941a) Animal

XVII

Nutritional Assessment of the Individual

JOHN YUDKIN

The approach to the problem of assessing nutritional state tends to be different for the nutritionist and for the clinician. The approach of the nutritionist is usually statistical; he is concerned in the assessment of nutritional state in groups or populations, comparing one group with another group or de- [illegible] various [illegible] is more [illegible] r whom a diagnosis is sought, and where nutritional deficiency may play a part in the pathogenesis of the signs and symptoms with which the patient presents himself. The purpose of this paper is to indicate that, though the guiding principles in diagnosis are the same for both nutritionist and clinician, the application and interpretation of these principles are different, and present particular problems to the clinician

THE DIAGNOSIS OF NUTRITIONAL DEFICIENCY

The diagnosis of nutritional deficiency is made by the application of three criteria (Yudkin, 1944a).

1. There should be evidence of an inadequate supply of one or more of the essential nutrients

2. There should be evidence that the signs and symptoms are those which are known to occur with deficiency of the suspected nutrients.

3. The correction of the deficiency—that is the administration of the suspected nutrients—should cure the condition.

The proof that nutritional deficiency is the cause of a disease thus depends on applying a series of principles, in the same sort of way that Koch's postulates are used to prove the implication of a micro-organism in the causation of a disease. However, in the same way that it is not always possible completely to demonstrate all three of Koch's postulates in microbial disease, so it is not always possible completely to satisfy all the diagnostic criteria in the diagnosis of nutritional deficiency. In both instances, it is sometimes necessary to depend on incomplete or circumstantial evidence, but if error is to be avoided, it is essential that all the possible evidence be sought and assessed before a decision is reached. [illegible]

claims for the existence and cure of nutritional deficiency.

The use of these criteria in particular instances and the conclusions drawn from the evidence require some care and understanding, as I have briefly described elsewhere (Yudkin, 1944a). One factor of particular importance, which is often overlooked in the field of nutrition, is that of variation—both variation between individuals and variation in one individual at different times. The existence of this variation adds considerably to the difficulty in the assessment of nutritional status in individuals, but we shall see that there are ways in which it may be allowed for in the assessment of groups.

The first diagnostic criterion is the determination of the adequacy or otherwise of the intake of all essential nutrients. To determine adequacy of intake, it is necessary not only to know the intake—what food and how much food, and so how much of each nutrient—but also to know the requirements of the individual for each nutrient. It is now becoming clear that it is not easy to determine either requirements or intakes with any degree of accuracy and that many of these difficulties are due to the variations mentioned earlier (Yudkin, 1948, 1951).

VARIATIONS IN NUTRITIONAL REQUIREMENTS

In considering the question of nutritional requirements, it is necessary first to decide what is meant by this term. Most

clinicians would now accept the view that there is a gap—often a wide gap—between the amount of a nutrient which will just prevent signs of obvious deficiency and the amount which will produce optimal nutrition. The amount of vitamin C which will just prevent scurvy is small; we might call this the minimal requirement. The larger amount required to give optimal nutrition, so that a further increase would produce no further improvement, would then be called the optimal requirement. An intake between the minimal and the optimal would lead to a state of sub-optimal nutrition. In practice, it is found that, whereas it is fairly easy to determine the minimal requirement, it is usually far more difficult to determine the optimal requirement.

This difficulty is due to a combination of two factors—the difficulty of assessing sub-optimal nutrition and the decreasing effect of increasing intakes. The difficulty of assessing sub-optimal nutrition derives from the fact that this state may show itself only in slightly impaired growth in children, or a decrease in subjective well-being, or a decreased resistance to infection, all of which may be difficult or impossible to detect in an individual. The second difficulty in detecting optimal nutrition is that illustrated in Fig. 1. When the intake of a nutrient is low—below the minimal requirement—progressive small increments produce relatively great and easily detectable improvement. When the intake is higher, these small increments may still be producing an improvement but only at a continuously reducing rate. In the example given in Fig. 1 three units of the nutrient prevent obvious deficiency disease: this is then the minimal requirement. But it is not so easy to see how many units produce the maximal response. It might be that the maximal response occurs with twelve units; it might however be that a further slight response occurs when the intake is increased to eighteen units. If so, this is a very large difference in intake producing only a very small difference in response. Moreover, not only is the difference small but, as we have seen, it may be in some quality which itself is difficult or impossible to measure.

Another difficulty in the determination of requirements is that the methods are difficult and laborious, and that they do not

always give the same results. One method is to study a population in which a deficiency disease is endemic and determine the intake of, say, vitamin B_1 in those who suffer from beri-beri and in those who are free from this disease. The same sort of method has been used experimentally, groups of volunteers being given diets containing graded doses of one nutrient and the amount ascertained which just prevents the onset of signs or symptoms.

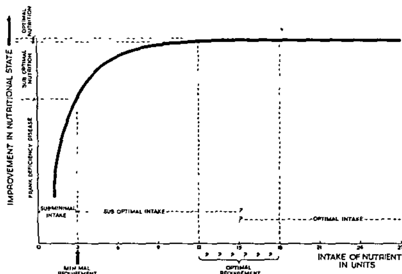


FIG 1. The effect of varying amounts of nutrient on nutritional state

Another method is based on so-called balance experiments, in which is determined the amount of, say, calcium which is needed just to prevent a net loss to the body. A further method is to find how much of a nutrient is needed to saturate the body, so that excess spills over into the urine. For some nutrients, none of these methods is practicable and assessment of requirements is based on the results of experiments with animals, the values being then calculated for human subjects Occasionally, it is only possible to assess average intakes and base requirements on the fact that these intakes represent something like the amounts normally needed. This may, however, be far higher than the real needs; for example, it appears that an intake of fat of 5–10

clinicians would now accept the view that there is a gap—often a wide gap—between the amount of a nutrient which will just prevent signs of obvious deficiency and the amount which will produce optimal nutrition. The amount of vitamin C which will just prevent scurvy is small; we might call this the minimal requirement. The larger amount required to give optimal nutrition, so that a further increase would produce no further improvement, would then be called the optimal requirement. An intake between the minimal and the optimal would lead to a state of sub-optimal nutrition. In practice, it is found that, whereas it is fairly easy to determine the minimal requirement, it is usually far more difficult to determine the optimal requirement.

This difficulty is due to a combination of two factors—the difficulty of assessing sub-optimal nutrition and the decreasing effect of increasing intakes. The difficulty of assessing sub-optimal nutrition derives from the fact that this state may show itself only in slightly impaired growth in children, or a decrease in subjective well-being, or a decreased resistance to infection, all of which may be difficult or impossible to detect in an individual. The second difficulty in detecting optimal nutrition is that illustrated in Fig. 1. When the intake of a nutrient is low—below the minimal requirement—progressive small increments produce relatively great and easily detectable improvement. When the intake is higher, these small increments may still be producing an improvement but only at a continuously reducing rate. In the example given in Fig. 1 three units of the nutrient prevent obvious deficiency disease this is then the minimal requirement. But it is not so easy to see how many units produce the maximal response. It might be that the maximal response occurs with twelve units; it might however be that a further slight response occurs when the intake is increased to eighteen units. If so, this is a very large difference in intake producing only a very small difference in response. Moreover, not only is the difference small but, as we have seen, it may be in some quality which itself is difficult or impossible to measure.

Another difficulty in the determination of requirements is that the methods are difficult and laborious, and that they do not

700 mg If our subject is consuming more than this, he is certainly getting enough. But if, as is more likely, he is consuming somewhat less than 700 mg, we can obviously not conclude that he is deficient in calcium. Even if he is consuming half of this amount, that is 350 mg daily, we know from our results that he may be one of that 20 per cent of the population for whom this is an adequate intake. This last statement provides a clue to the only legitimate way in which we can assess the adequacy of intake compared to requirement. Provided we know the distribution of the values for the requirement of a nutrient, we can express the adequacy of intake as a probability. Thus, a person whose daily intake of calcium is 450 mg has an even chance that he is getting enough; one with an intake of 600 mg has a 9 to 1 chance that he is getting enough, one with an intake of 350 mg has a 4 to 1 chance that he is *not* getting enough.

As has been said, the fact that these variations in requirements exist from person to person is only another example of the biological variation which is known to exist between individuals of a species. It is, however, possible to go a little further in explaining the variations in nutritional requirements. First, we are beginning to realize that the requirements for some nutrients are dependent on the other components of the diet. The dependence of the calcium requirement on the content of dietary phytic acid is well known. Again, the requirement of nicotinic acid is higher when the diet contains maize, since this food appears to contain an antivitamin which interferes with the utilization of nicotinic acid. In addition, the amino-acid tryptophane can replace nicotinic acid in the diet, so that the requirements of the vitamin are lower when the diet contains more tryptophane As a result of this knowledge, we can now say that the association of pellagra with the consumption of maize diets may be explained by the low content of nicotinic acid, the low content of tryptophane and the presence of an antivitamin As another example of the dependence of requirements upon other dietary constituents, we may cite the increased requirement of some vitamins of the B group when the intake of others is increased, and the decreased requirement of vitamin A when the intake of vitamin E is increased. There is also an alteration in require-

grams daily is quite adequate, yet this would, in Western dietaries, be something like one-tenth of the amounts normally consumed and would in fact be quite incompatible with our dietary habits.

Apart from the tedious nature of some of these methods, the results given by one method are often very different from those given by another method. With vitamin C, an intake of about 10 mg daily is enough to prevent scurvy, but something like 100 mg daily are needed for saturation. Even if it is accepted that the body requires more than just enough to prevent scurvy, the gap between the two is so large that it is difficult to know what figure really represents the needs of the body. This difficulty has still not been resolved, for British workers are inclined to accept, for a normal man, a value of 20 or 30 mg, whilst American workers accept a value of 75 mg (B.M.A., 1950; N.R.C., 1948).

Nevertheless, let us for a moment assume, which is in fact far from true, that it is possible to determine with some degree of accuracy the requirements for each nutrient for a series of individuals. We find at once, as indeed might have been expected from our general biological knowledge, that the requirements are not the same for each person If we measured, as an index of calcium requirement, the minimal amount of calcium that each one of a series of 100 persons requires to produce a balance between intake and output, we should probably find that, on average, this is 450 mg a day. But of the 100 persons studied, 20 might require less than 350 mg and 10 as little as 300 mg; similarly, 20 might require 550 mg, 10 more than 600 mg and one or two as much as 700 mg

The implications of these variations between individuals are important. We may, from a knowledge of the extent of the variation, adopt either of two values for the requirement of calcium. We may adopt the value of 450 mg on the justifiable grounds that it is the average. If we do this, however, we have no means of knowing whether a particular person consuming just this amount belongs to that large [illegible] s whose requirements are [illegible]

Alternatively, we ma [illegible] requirement of

VARIATIONS IN INTAKE

It is clear then that it is not at all easy to establish exactly the requirement of any one nutrient for any one person. Even if it were possible, however, it would be necessary to establish fairly accurately what the intake of the nutrients is in order to be able to determine whether any one of these is inadequate. Investigation has shown that it is in fact not at all easy to determine dietary intake accurately. Apart from technical difficulties, which are quite substantial, there appears to be considerable variation in intake over a period of time. The fact that the intake of food and hence of nutrients varies from day to day is well known and not so important. It is, however, assumed that, at least in civilized countries, the intake over a week is representative of the usual intake except perhaps for seasonal differences in the availability of some foods. In fact, it is found that a person may consume quite different types and amounts of food from week to week. The extent of the variation itself varies not only from person to person but from nutrient to nutrient. Of six subjects studied over four consecutive weeks in a recent investigation, the caloric intake of one was almost constant, whilst that of another was nearly 70 per cent more in one week than in another (Fig. 2). Of iron, one subject consumed a daily average of as little as 7 mg in one week and as much as 14 mg in another week (Fig 3). Of vitamin C, another subject consumed daily 45 mg in one week and 109 mg in another (Fig. 4). The variation was greatest with vitamin A and vitamin D, with as much as a fivefold and an eightfold difference between the weeks of lowest and highest consumption (Fig. 5). The most important finding in this investigation was the fact that, both with calories and with nearly every one of the nutrients studied, at least one subject would be considered to have had a deficient intake in one week and an adequate intake in another week, when compared with any of the commonly accepted standards of requirements. A study based on one week's intake, and still more a study based on a few days' intake, is thus quite likely to give an erroneous assessment of the true intake and more particularly an erroneous assessment of the nutritional adequacy of the diet.

ment of some of the vitamins in the B group with variation in the intake of fat, carbohydrate and protein.

A second possible cause of variation in requirements between individuals is related to fluctuations which may occur from time to time in one individual. We know the causes of some of these fluctuations; there is an increased need for some vitamins of the B group with increased exercise, and of these and many other nutrients in pregnancy and lactation. Increased metabolism in hyperthyroidism or fever may also increase the need for some nutrients. These factors and no doubt others that we do not yet understand mean that, in a group of persons at any one time, part of the differences in individual requirements is due to differences in their physiological as well as in their pathological state.

Another possible cause of variation may lie in the fact that some nutrients, particularly some of the vitamins of the B group, can be synthesized by intestinal micro-organisms. The *dietary* requirements for these nutrients, then, must be the difference between the actual needs by the tissues, and the amount synthesized by microbial action. Differences between persons in the amounts synthesized, which would cause differences in dietary requirements, might be due amongst other things to differences in the anatomical and physiological conditions in the intestinal tract.

We need only consider one more possible cause of individual variations in requirements. This is the possibility that a person may become adapted to a high or to a low intake of a nutrient. There is some evidence that persons may require less than the normal amounts of vitamin B_1 or of calcium if they have subsisted for a time on a diet low in either of these substances. As a corollary, it would follow that, on a diet very restricted in vitamin B_1, a person who had been accustomed to a relatively low intake might develop beri-beri less easily than one who has been accustomed to a higher intake. We know as yet little about this phenomenon, but it might well provide one explanation for differences in requirements between individuals.

VARIATIONS IN INTAKE

It is clear then that it is not at all easy to establish exactly the requirement of any one nutrient for any one person. Even if it were possible, however, it would be necessary to establish fairly accurately what the intake of the nutrients is in order to be able to determine whether any one of these is inadequate. Investigation has shown that it is in fact not at all easy to determine dietary intake accurately. Apart from technical difficulties, which are quite substantial, there appears to be considerable variation in intake over a period of time. The fact that the intake of food and hence of nutrients varies from day to day is well known and not so important. It is, however, assumed that, at least in civilized countries, the intake over a week is representative of the usual intake except perhaps for seasonal differences in the availability of some foods. In fact, it is found that a person may consume quite different types and amounts of food from week to week The extent of the variation itself varies not only from person to person but from nutrient to nutrient. Of six subjects studied over four consecutive weeks in a recent investigation, the caloric intake of one was almost constant, whilst that of another was nearly 70 per cent more in one week than in another (Fig. 2) Of iron, one subject consumed a daily average of as little as 7 mg in one week and as much as 14 mg in another week (Fig 3). Of vitamin C, another subject consumed daily 45 mg in one week and 109 mg in another (Fig. 4). The variation was greatest with vitamin A and vitamin D, with as much as a fivefold and an eightfold difference between the weeks of lowest and highest consumption (Fig. 5). The most important finding in this investigation was the fact that, both with calories and with nearly every one of the nutrients studied, at least one subject would be considered to have had a deficient intake in one week and an adequate intake in another week, when compared with any of the commonly accepted standards of requirements. A study based on one week's intake, and still more a study based on a few days' intake, is thus quite likely to give an erroneous assessment of the true intake and more particularly an erroneous assessment of the nutritional adequacy of the diet.

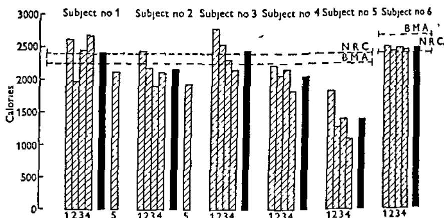


FIG. 2. Variation in weekly intake of *calories* of six subjects

(or five) weeks of consecutive weeks. Research Council

(1948) and British Medical Association (1950).

FIG. 3 Variation in weekly intake of *iron* of six subjects. For further information, see Fig 2

Vitamin C (mg)
100
90
80
70
60
50
40
30
Subject no 1
Subject no 2 Subject no 3 Subject no 4 Subject no 5
Subject no 6
NRC
BMA

FIG. 4. Variation in weekly intake of *vitamin C*.
For further information, see Fig 2.

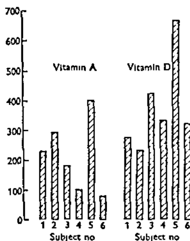


FIG. 5. Differences between highest and lowest weekly intakes of vitamin A and vitamin D of six subjects. Figures give difference as percentage of lowest intake.

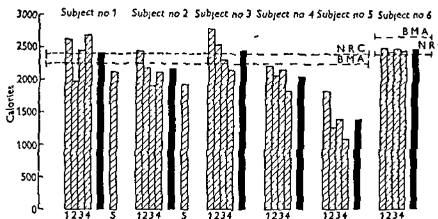


Fig. 2. Variation in weekly intake of *calories* of six subjects.

... average daily intake for four (or five) weeks of ... re weeks, ... Council

Subject no 1 Subject no 2 Subject no 3 Subject no 4 Subject no 5 Subject no 6

Iron (mg)

20 15 10 5 0

BMA NRC

1234 5

Fig. 3 Variation in weekly intake of *iron* of six subjects.
For further information, see Fig 2

Vitamin C (mg)

100 90 80 70 60 50 40 30

Subject no 1 Subject no 2 Subject no 3 Subject no 4 Subject no 5 Subject no 6

N R C

B M A

FIG. 4. Variation in weekly intake of *vitamin C.*
For further information, see Fig. 2.

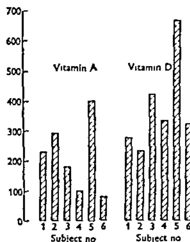


FIG. 5. Differences between highest and lowest weekly intakes of vitamin A and vitamin D of six subjects. Figures give difference as percentage of lowest intake.

From the discussion so far, it is evident that it is not possible to determine for one person exactly how much of each nutrient he requires, nor to determine exactly how much of each nutrient he normally consumes. In spite of this, in many instances even a rough assessment of intake will show that the individual is taking quite enough or alternatively quite insufficient of one or more nutrients. A person who eats meat twice a day, and whose diet in addition contains reasonable quantities of eggs, milk and cheese, is most likely to be taking enough protein, whatever we accept as the protein requirement. Conversely, the diet consumed by an Indian peasant, as revealed by a dietary history and a most superficial measurement of the foods consumed, will readily reveal deficiencies, even if we accept quite low standards of requirements. Nevertheless, the diets of many persons will be within the range in which the uncertainties in determining intake and requirement make it impossible to decide whether or not there exists some degree of deficiency.

VARIATIONS IN BODILY FUNCTION AND COMPOSITION

As another example of the effect of variability on the diagnosis of nutritional deficiency, we may consider the use of the signs revealed by clinical and laboratory examination. It is generally accepted that clinical signs, in any disease, may differ greatly from person to person But this tends to be forgotten with the more specialized tests used for the diagnosis of nutritional deficiency and particularly with laboratory tests which involve, for example, estimation of nutrients in body fluids. As a result, a great deal of unwarranted conclusions have been published regarding the incidence of deficiency. One example is the alleged deficiency of riboflavin It will be recalled that one of the signs of this condition is an increased vascularity of the limbal region of the cornea. From the results of an estimation based entirely on this sign, it was suggested that deficiency of riboflavin occurred in a large proportion of apparently normal people—for example, in over 75 per cent of New York school children (Wiehl and Kruse, 1941) Apart from the fact that an increase in corneal vascularity can be caused by other factors

than deficiency of riboflavin, it is now evident that studies such as these ignored entirely the normal range of variation in the number and complexity of the vessels found in this region of the eye. Similar errors were made in the interpretation of measurements of dark adaptation, where a standard of normality was set up by many workers and performance below this standard was taken as indicative of deficiency of vitamin A (for example, Jeans and Zentmire, 1936; Maitra and Harris, 1937). Apart from the fact that there are other causes of impaired dark adaptation than deficiency of vitamin A, there is a deterioration of dark adaptation with age (Robertson and Yudkin, 1944). Moreover, it is now quite clear that, at any one age, there is not one normal value but a range of normal values

If it is accepted that there is a normal range rather than a normal value in these and similar instances, we are led to the same sort of conclusions as we derived in considering dietary requirements. It is possible, that is, only to assess the probability of normality from the results of examinations. Thus, a person whose dark adaptation is within the range of normal may in fact be normal, or may on the other hand have some mild pathological condition, for example deficiency of vitamin A, which has brought the value down from a somewhat better standard. The nearer the value is to the best in the range, the higher the chances that no pathological state exists; the nearer it is to the worst in the range, the lower the chance that no pathological state exists. Only if the value is equal to the best in the range, or inferior to the worst, can a definite conclusion as to normality or abnormality be reached.

It might perhaps be worth while to point out that whilst these considerations apply to the nutritional assessment in an individual, somewhat different considerations apply when groups of subjects are to be compared. With groups it is possible to compare relative nutritional status by comparing the mean values, or the distribution of values (Fig 6) If it can be taken that two groups are similar except as regards their diet, for example that they are of similar age and sex, then an inferior dark adaptation in one group is likely to be due to a lower intake of vitamin A, or a lower level of haemoglobin to a deficiency of iron, or a lower

height and weight of children to a general nutritional deficiency (Yudkin, 1944b, 1945). *But it is still impossible to say from these* results which individuals in the groups are deficient.

The importance of realizing the existence of a range—even a fairly wide range—of normal values in the body constituents is confirmed by some recent work on pallor and haemoglobin

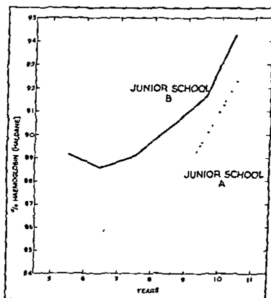


FIG 6 Haemoglobin levels in school children. Children from School A were of lower economic status than those from School B.

levels in children. Many clinicians are aware that pallor is not necessarily a sign of anaemia. Yet many are still inclined to treat a patient for anaemia, by administration of iron either alone or in come complex mixture with other minerals, or even by parenteral administration of liver, solely on the finding that the patient looks pale. An analysis of a comprehensive survey of about 1,200 school children has shown quite definitely that there is no association whatever between pallor and haemoglobin values (Yudkin, 1952). The children had a range of haemoglobin from about 75 per cent Haldane to over 100 per cent Haldane. They were apparently healthy so that these results do

not mean that real anaemia, with haemoglobin levels of say 50 per cent, is not associated with pallor.

Further analysis of the observations in the survey showed that the level of haemoglobin was not correlated with any of the numerous indices which are usually used as criteria of nutritional status in children. Thus, there was no correlation with height or weight or with the Tuxford index, nor with clinical assessment of nutrition. A careful consideration of all the findings makes it difficult to resist the conclusion that haemoglobin levels appreciably below the accepted minimal values are compatible with normal physiological function and that in fact there is a fairly wide normal range of haemoglobin, rather than a single normal value. Whilst it may be surprising at first to be asked to agree that a haemoglobin level some 75 per cent of the usually accepted standard may still be a normal value, it is in fact only another example of variability in bodily constitution.

CORRECTION OF THE DEFICIENCY

There is one important method by which it is possible in the diagnosis of nutritional deficiency to overcome the difficulty created by the wide range of normal variations This is to apply the therapeutic test, the third of the diagnostic criteria mentioned at the beginning of this paper. By means of this test, it is often possible to detect whether a person is deficient or whether he is within the lower part of the range of normal. This is the basis, for example, of the use of the so-called 'vitamin A labile threshold' in detecting deficiency of this vitamin. The principle is to measure dark adaptation both before and after administration of vitamin A (Yudkin, Robertson and Yudkin, 1943). A significant improvement in dark adaptation following the administration is reasonable evidence that deficiency has existed.

Two points must, however, be borne in mind. First, even though a sign may be due to a specific nutritional deficiency, it may have become irreversible so that correction of the deficiency may not effect a cure. This is true of some of the neurological changes found in beri-beri. Second, administration of a nutrient may produce improvement in a clinical condition which was not, however, caused by nutritional deficiency. An example is

the effect of vitamin D in lupus vulgaris, a condition which is certainly not due to avitaminosis D. Here and in other instances, the nutrient is presumably acting as a pharmacological agent rather than in the way of making good a deficiency.

CONCLUSION

It may be concluded that diagnosis of nutritional deficiency in the individual must take into account the wide variation from time to time in what we might call the external environment, that is in the dietary intake, and in the wide variation from person to person in the internal environment, that is in their requirements and in the composition and the functioning of their bodily tissues. Nevertheless, if care is taken to adduce evidence from as many directions as possible, and use is made of all three diagnostic criteria, it is often possible to achieve a satisfactory nutritional assessment of the individual.

REFERENCES

B.M A (1950). *Report of the Committee on Nutrition.* British Medical Association, London

JEANS, P. C., and ZENTMIRE, Z (1936) *J Amer. med. Ass.* **106**, 996.

MAITRA, M K, and HARRIS, L J (1937) *Lancet*, **2**, 1009

N.R C (1948) *Rep nat Res. Coun, Wash.*, No. 129

ROBERTSON, G W, and YUDKIN, J (1944) *J. Physiol* **103**, 1.

WIEHL, D. G, and KRUSE, H D (1941) *Milbank Mem Fd quart Bull* **1**), 241

YUDKIN, J. (1944a) *Brit Med. J* **1**, 5

— (1944b). *Brit med J* **2**, 201.

— (1945) *Proc roy Soc Med* **38**, 162

— (1948) *Nutr Diet Cater* **2**, 43

— (1951). *Brit J Nutr* **5**, 177

— (1952) *Lancet*, **1**, 239.

— ROBERTSON, G W, and YUDKIN, S (1943) *Lancet*, **2**, 10.

XVIII

Biological Action of Radiation

J. F. LOUTIT

BECQUEREL (1896) discovered that pitchblende emitted radiations that fogged photographic plates. This was the first demonstration of natural radioactivity. After the Curies had isolated radium from pitchblende, Becquerel carried for demonstration purposes a tube of radium in his waistcoat pocket. It is reported that, while it did not burn a hole in his pocket, it produced a burn in his underlying skin. This is not the earliest incident of the biological effect of ionizing radiation but it is a historic one. The discovery of X-rays by Rontgen (1895) and their biological effects was made about the same time. As X-ray tubes in those early days emitted radiations of comparatively low penetrative power and as radioactive sources were small in activity, the accidental destructive lesions resulting from handling them at first tended to be local. It was clear that attempts would be made to use this biological effect in a controlled fashion in medical therapy for unwanted tumours and the like. In later years when radiologists and radium-chemists began to die from systemic diseases it was realized that exposure of the whole body to radiation or ingestion of radioactive substances produced a generalized deleterious effect, either as a refractory anaemia or through the induction of malignancies. However these incidents were sporadic and though they created an occupational hazard, which was gradually overcome by increasing awareness of the need for personal protection, most of the investigations into the biological action of radiations were made on local tissues, isolated tissue cultures or on suspensions of unicellular organisms. These fundamental studies have been

admirably reviewed by Lea (1946), Spear (1946), Duggar (1936), etc.

With the use of the atomic bomb the problem of the clinical effects of energetic radiation delivered to the whole body has been once again pushed to the fore. I have chosen this subject to discuss today. In the time available it is possible to discuss only one facet of the biological effects of radiation and I do not mean to imply that research is complete as regards the fundamentals or that applied research is not being pursued as actively as hitherto in the treatment of cancer and similar diseases. On the contrary with such rapid progress in atomic physics the possibilities are brighter and the scope greater than ever before. Moreover, while the effects of irradiation with β-rays and neutrons may be relevant, I will confine discussion today to the effects of γ-rays or X-rays within the deep-therapy range of 200 kV upwards.

As with any damaging agent the effects of penetrating X- or γ-rays can be lethal if the dose applied to the whole body of an animal is sufficiently large. Therefore the standard pharmacological practice of determining dose-effect curves and dose-time of effect curves is applicable. Within the mammalian order, there is remarkably little variation between the genera. Doses of X- and γ-radiation are measured in units of rontgens.[1] The dog would seem to be about the most radiosensitive mammal with a MLD of *circa* 300–350r (Field and Rekers, 1949; Boche and Bishop, 1946) and the rabbit the most radioresistant at *circa* 800r, with the mouse, monkey, rat and hamster intermediate (Boche and Bishop). These American authors quote the guinea-pig as being even more sensitive than the dog with a MLD of 200r: in my laboratory, however, the figure is nearer 500r for hybrid animals and at the National Cancer Institute, Bethesda, 400r for inbred strains (Lorenz *et al.*, 1951) A variation between laboratories with different methods of irradiation, different methods of measurement and different stocks is not surprising. The MLD for man has been assessed from the Japanese atomic

[1] One rontgen is that quantity of X- or γ-radiation the associated corpuscular emission of which in 0 001293 gm of air gives one electrostatic unit of electricity of either sign

bomb incidents as around 400r (U.S. Atomic Energy Commission and Department of Defense, 1950). In the experimental animals the deaths from the acute effects of radiation occur within three weeks: in man within about six weeks.

THE MEDIAN LETHAL DOSE OF RADIATION

The 'radiation syndrome' which results in man from a dose of radiation in the region of the median lethal is divisible into three stages·

(1) An initial phase of '*radiation sickness*'. The term 'radiation sickness' has long been in use for that clinical upset which follows extensive deep radiotherapy especially of the abdomen and some say the neck. Within a few hours of the irradiation nausea, vomiting and loss of appetite ensue. Associated with this there is frequently a state of 'shock'. Good descriptions of this state have not appeared in the recent literature. There is certainly a profound physical lassitude or prostration with some dizziness and mental irritability. Cardiovascular signs with fall in blood pressure have been reported but are certainly far from constant. The clinical 'shock-like' state is not reminiscent of oligaemic 'shock' from haemorrhage.

Most experimental animals when given lethal doses do not exhibit this immediate effect to the same degree. The dog, it is true, may vomit and show some effect. Rabbits occasionally die suddenly and unexpectedly, possibly from cerebral anoxia; Prosser (1947) reports a profound fall in blood pressure after lethal doses of radiation. Most animals seem to be less active and to take less food than usual for a day or two. But when seeking evidence for oligaemia by determinations of the plasma volume of rabbits with Evans blue, Painter *et al.* (1946) reported normal

It is known from histological evidence (Bloom, 1948, and earlier authors) that within a matter of hours of such a large dose of radiation there is in the normal experimental animal a widespread destruction of the small and medium lymphocytes and the beginnings of an almost complete loss of the red bone

marrow. Estimates of the mass of the red marrow and lymphatic tissue suggest that in man this may amount to 2 kg or more (International Prot. Com. (quoting Lisco), 1951). This rapid dissolution of tissue could easily be a source of toxic products of protein or H-substances (Barnes and Furth, 1943; Caspari, 1928; Beclere, 1918).

Certainly in the rabbit Painter *et al.* (1946) found that cross-transfusion of the shocked irradiated hypotensive rabbit with a normal rabbit caused a fall in blood pressure in the latter animal. They also claimed to find an increased content of plasma histamine in the early post-irradiation period and so did Segal (1939) in human patients after clinical radiotherapy. Ellinger (1945, 1946), another protagonist of this hypothesis involving histamine, links it with stress and the adaptation syndrome of Selye. He quotes Leblond and Segal (1942) that the fatty infiltration of the liver seen both after histamine injections and after irradiation is not seen in irradiated adrenalectomized animals.

Although the plasma volume appears to be unchanged, the water balance of the body is disturbed and permeability of membranes is upset. After local irradiation of the skin of experimental animals, the irradiated areas become dyed when intravenous injections are made of pyrrol blue (Mottram, 1933) and trypan blue (Rigdon and Curl, 1943). It is further claimed that tagged protein atoms leave the circulation more rapidly than normal (Storey *et al*, 1952). Experimental animals at this stage lose weight rapidly. Much of this is due to failure to eat and drink, but there is also an increased output of urine, certainly in the rabbit (Prosser *et al.*, 1947) and in the rat, which has been attributed to diminution in output of antidiuretic hormone (Edelmann and Eversole, 1951).

Because 'radiation sickness' has been an undesirable complication of radiotherapy numerous attempts have been made to combat it or prevent it The method of approach has usually been the therapeutic trial to test a particular hypothesis. In the published reports the trial has almost invariably been successful. When such diverse substances as an antihistamine (Lofstrom and Nurnberger, 1946) and histaminase (Ellis, 1942); pressor agents (Jenkinson and Brown, 1944); desoxycorticosterone ace-

tate (Ellinger *et al.*, 1949); thiamine and nicotinic acid (Bean *et al.*, 1944; Imler and Wammock, 1940) and pyridoxine (Shorvon, 1949) are all highly recommended it suggests that none is specific and that each may be in its way countering some part of a widespread disturbance.

(2) *A latent interval.* Following the initial stage with its multiple subjective symptoms lasting up to several days, comes a stage of comparative subjective well-being. In man this may last several weeks but in the experimental animal is rarely longer than 1½–2 weeks. According to dose-time studies the period is shorter the greater the dose received.

Objectively the most noticeable sign is the leucopenia of the peripheral blood. From the time of the irradiation, changes in the leucocyte count are noted. Reliable data do not exist for man In the Hiroshima and Nagasaki incidents there was too great a chaos for such data to be collected in the first few days. It was not till the arrival of the U.S. forces at the third week that haematological investigations were made on a large scale. These have been reported by Le Roy (1950) From the earliest days of radioactivity investigations have been made with experimental animals. Probably the most extensive is that on rabbits by Jacobson *et al.* (1947). The studies confirmed the rapid and almost complete loss of circulating lymphocytes. Within a matter of a few hours following a mid-lethal dose these cells are reduced to about 10 per cent of the normal figure and remain at or around this level for many days. The granular leucocytes (heterophils) within the first day or so were increased in number. In Jacobson's animals given high doses, two maxima were observed at 8 and 24 hours. In my own laboratory when counts were made at more frequent intervals in the early hours after irradiation, but not overnight, an additional maximum was found at 1½ hours (Fig. 1) (Barnes, 1952). Obviously the mature granulocytes and band cells in the bone marrow are not completely destroyed like their precursors and can be delivered to the circulation in response to the stress of the irradiation and its sequelae. The shape of the curve of the heterophil count in Fig. 1 might suggest that there were successive waves of delivery from the marrow to the peripheral blood corresponding to the

marrow. Estimates of the mass of the red marrow and lymphatic tissue suggest that in man this may amount to 2 kg or more (International Prot. Com. (quoting Lisco), 1951). This rapid dissolution of tissue could easily be a source of toxic products of protein or H-substances (Barnes and Furth, 1943; Caspari, 1928; Beclere, 1918).

Certainly in the rabbit Painter *et al.* (1946) found that cross-transfusion of the shocked irradiated hypotensive rabbit with a normal rabbit caused a fall in blood pressure in the latter animal. They also claimed to find an increased content of plasma histamine in the early post-irradiation period and so did Segal (1939) in human patients after clinical radiotherapy. Ellinger (1945, 1946), another protagonist of this hypothesis involving histamine, links it with stress and the adaptation syndrome of Selye. He quotes Leblond and Segal (1942) that the fatty infiltration of the liver seen both after histamine injections and after irradiation is not seen in irradiated adrenalectomized animals.

Although the plasma volume appears to be unchanged, the *water balance of the body is disturbed and permeability of* membranes is upset. After local irradiation of the skin of experimental animals, the irradiated areas become dyed when intra-

normal (Storey *et al.*, 1952). Experimental animals at this stage lose weight rapidly. Much of this is due to failure to eat and drink, but there is also an increased output of urine, certainly in the rabbit (Prosser *et al.*, 1947) and in the rat, which has been attributed to diminution in output of antidiuretic hormone (Edelmann and Eversole, 1951).

Because 'radiation sickness' has been an undesirable complication of radiotherapy numerous attempts have been made to combat it or prevent it. The method of approach has usually been the therapeutic trial to test a particular hypothesis In the *published reports the trial has almost invariably been successful.* When such diverse substances as an antihistamine (Lofstrom and Nurnberger, 1946) and histaminase (Ellis, 1942), pressor agents (Jenkinson and Brown, 1944); desoxycorticosterone ace-

after irradiation a significant percentage of circulating lymphocytes show cytological signs of degeneration. These degenerating cells may have come, however, from the lymph depôts.

The heterophil leucocytes of the rabbit disappear rapidly from the circulating blood after the first day. By the fourth to sixth day both heterophils and lymphocytes are at a minimum of about 10 per cent or less of the normal count. The heterophils

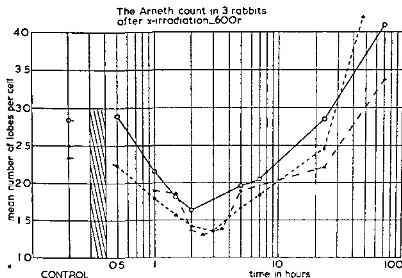


Fig 2

begin to increase again towards the end of the latent period, but the recovery of the lymphocytes is much slower. The pattern of loss and recovery of the different categories of circulating leucocytes almost certainly differs from one species of animal to another Preliminary studies (Hunter, 1952) in our laboratory with monkeys (Macacus rhesus) confirm this. The pattern of recovery at least in the experimental rabbit is quite similar to that of the Japanese human as reported by Le Roy (1950)

The circulating elements other than the leucocytes are also affected by the irradiation After the leucocytes the platelets are the next to be grossly depleted. In Le Roy's cases the low levels persisted for many weeks and, following radiotherapy, Court-

peaks, or that there were successive waves of elimination of the circulating cells corresponding to the troughs: it could also be that circulating cells were periodically sequestered, as in lung (cf. Weisberger *et al.*, 1950), and released again. Arneth counts performed on a number of these animals showed a progressive fall in the average number of lobes per cell from time zero to

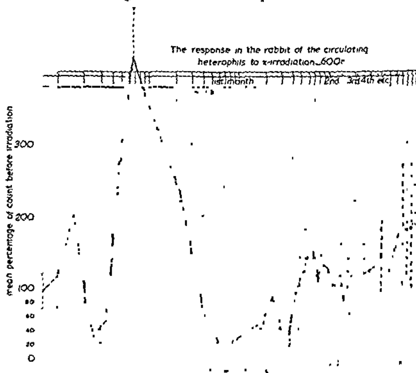


Fig. 1

4 hours and then a progressive rise (Fig. 2) This finding does not favour either the hypothesis of periodic release from the marrow or sequestration and release. The combined data of Figs. 1 and 2 show that a considerable proportion of the circulating heterophils are removed from the circulation in the first *two hours*. It has been generally supposed that the circulating leucocytes are not affected by doses of radiation of this order, since in blood irradiated *in vitro* the leucocytes are apparently unaffected. Trowell (1952) has shown that in the rat a few hours

(*a*) *Sepsis and septicaemia.* Blood cultures on dogs (Bennett *et al*, 1951) irradiated at the midlethal level showed that bacteriaemia occurred with moderate frequency, and in animals dying late in first few post-irradiation weeks a positive blood culture was the rule. The gram-negative organisms, the normal inhabitants of the intestinal tract, seemed to be the most significant. Similarly Miller *et al.* (1950a) found a high rate of positive blood and spleen cultures in irradiated mice and attributed their bacteriaemia to invasion from an intestinal reservoir. It was logical therefore to investigate the possible prophylactic effects of the new chemotherapeutic drugs and antibiotics. Cronkite (1949) reported, however, failure to influence the death rate in mice by producing intestinal bacteriostasis with sulphonamides. Howland *et al.* (1949) using rats were able to prevent the terminal diarrhoea following a LD_{90} of X-rays, and to prolong life by a few days with aureomycin. The same team (Furth *et al.*, 1952) made similar findings in dogs. In both species of animal although the intestinal condition seemed to be partially controlled the animals died apparently from the haemorrhagic diathesis. Other antibiotics have also been tried in the same laboratory with similar or inferior results. On the other hand Miller *et al.* (1950b) found that streptomycin increased the survival of irradiated Swiss mice.

(*b*) *Haemorrhage.* The haemorrhagic diathesis has long been attributed to the thrombopenic element in the general picture of aplastic anaemia In 1948 the radiobiological field was shaken when Allen and his co-workers announced that the thrombopenia was relatively unimportant and that in the experimental dog the presence of a circulating anticoagulant was the major factor precipitating the haemorrhage They claimed that the effects of this factor, which resembled heparin, could be controlled by the antiheparins, protamine and toluidine blue. Animals so treated were not saved from death, but did survive longer and did not die from haemorrhage. About the same time another school produced evidence that the haemorrhage was attributable mainly to widespread vascular damage that could be prevented up to a point by premedication with flavone-glucosides, e.g. rutin (Griffith, 1947; Field and Rekers,

Brown (1948) believes the effect on the platelet count to be a more reliable indicator of over-dosage than that on the leucocytes. In the experimental animal (rat) Lawrence *et al.* (1948) in a review showed a maximum depression between 100 and 500 hours after the irradiation and a longer lasting depletion of megakaryocytes from the marrow.

The red cells, presumably because of their much longer natural mean life in the circulation, are diminished much more slowly. In man, where the mean life is 120 days or so, a decline in the red cell count of less than 1 per cent per day would be expected if erythropoiesis were completely suppressed. However an haemolytic as well as an aplastic factor has been cited (Schwartz *et al.*, 1947) and in the rabbit at least there is an appreciable drop in the total red cell volume as measured with ^{32}P and a concomitant fall in plasma volume for 10 days after which the plasma volume recovers while the red cell volume continues to decline (Storey *et al.*, 1950). Thus red cell counts and haemoglobin values give a rather distorted picture.

(3) *Delayed phenomena.* Because of the leucopenia one would conclude that the irradiated subject was highly susceptible to bacterial infection and this appears to be the case. Agranulocytic angina, ulcerative conditions of the stomach and intestine, necrotizing pneumonia and indolent skin ulcers were major manifestations in the Japanese cases (Liebow *et al.*, 1949). The histopathology of the lesions as described by these authors was notable for the intense invasion of the tissues by bacteria and almost complete lack of the usual tissue response of inflammation. Thus necrosis was the predominant feature.

Similarly because of the thrombopenia one would expect haemorrhagic lesions and these also occurred. The haemorrhages were distributed at random—as in purpura of the skin or urinary tract—or associated with the necrotic lesions. These haemorrhages led to an exacerbation of the anaemia. It is doubtful if death of such cases can be attributed to anaemia itself and anoxia, but the shock from haemorrhage in a debilitated subject or the toxaemia from bacterial infection or both could possibly account for it. A considerable amount of work on experimental animals has been conducted recently in the United States to test these hypotheses.

could be due to an alteration in permeability of membranes in general or to a failure of the hormonal regulatory mechanisms. Bowers and Scott (1951) report preliminary experiments showing that in rats the radiosensitive tissues lose potassium and that there is an overall potassium loss from the body: the movements of water and sodium were more difficult to interpret though sodium was on the whole strongly retained. Such phenomena do not necessarily indicate a direct effect on the 'sodium-pump'. On the other hand radiation is certainly a form of stress and effects from the adrenals are readily demonstrable. The adrenalectomized mouse is more sensitive to radiation than the normal (Cronkite and Chapman, 1950) and shielding of the adrenals with lead diminishes sensitivity (Edelmann, 1951). Pretreatment

cholesterol content of the adrenal followed by a considerable rise in both cholesterol and weight (Patt *et al.*, 1947). Detailed balances on the adrenalectomized or hypophysectomized animal do not seem to have been performed, but would certainly be in order.

Whereas these essays into balances and into the effects of the various hormones have not been exclusively performed on the animal given supralethal doses, the supralethal dose has been used widely for exploring the effects of potentially protective agents or measures.

Protective chemical agents. A large number of such agents are

shifts the LD_{50} of mice by about 10 per cent. Similar or rather greater effects have been obtained with cysteine but not cystine

priophenone (Storer and Coon, 1950); sodium nitrite (Cole *et al.*, 1950); sodium cyanide, sodium azide and malonitrile (Bacq

1949). This protective effect could not be elicited in mice by Cronkite *et al.* (1948). Cronkite and his collaborators (Penick *et al.*, 1951; Cronkite *et al.*, 1952) have recently been accumulating evidence in favour of the original hypothesis that the clotting defect and haemorrhagic tendency are due largely to the thrombopenia and can be corrected by transfusions of platelets.

In spite of these advances in the two fields there remains the strong suspicion that in addition to these factors there is a further biochemical lesion which is still unelucidated. Until it is, it is doubtful if appreciable progress will be made in the prevention or correction of this state which is undoubtedly strongly catabolic. Pyrexia or hyperpyrexia and a markedly negative nitrogen balance (Prosser, 1947) are indicative of this. It is probably the predominant factor when supralethal doses of radiation are considered.

SUPRALETHAL DOSE OF RADIATION

The dose mortality relationships with most mammals not strongly inbred are usually such that the lethal range extends over several hundreds of röntgens. Even with some inbred strains this is also seen, so that with a MLD of 500r the LD_{95} may be 800r. The supralethal range LD_{95} and upwards is regarded by many as a more useful one to investigate than the median. In the previous section it was noted that the survival time of the irradiated animal is inversely proportional to the dose received. After a supralethal dose therefore the experimental animal will succumb within a few days to a week, without signs of haemorrhage or overwhelming infection. It is true there is still the grossly leucopenic blood picture, but by itself this can scarcely be a killer and biochemical causes must be invoked.

The rat given 1,000r if irradiated with 240 kV X-rays from the dorsal aspect will die in about a week with signs of early haemorrhage: if irradiated from the ventral aspect it will die in a few days apparently from fulminating diarrhoea and dehydration. This can be obviated up to a point by treating the animals with subcutaneous saline (Mole, personal communication). This suggests that a major disturbance of the salt and water balances is a potent lethal factor at this time. The disturbance

red cells faster after irradiation than controls, and Cronkite *et al.* (1950) report that previous small 'conditioning' doses of X-rays protect mice against a later large dose.

The most notable effects are claimed by Jacobson *et al.* (1949 and 1951) from shielding with lead the exteriorized spleen of irradiated mice. The LD_{50} is raised from 550r to 1,100r. Even more impressive is the report that 50 per cent of irradiated mice can be induced to survive 1,025r if, after the irradiation, four spleens from infant mice are inserted into the peritoneal cavity. These authors favour a hypothesis that the shielded spleen or the implanted spleens liberate a hormone which hastens regeneration of the haemopoietic tissues rather than a seeding or colonization of cells from these unirradiated tissues The most potent factor in support of this is the demonstration by Lorenz *et al.* (1952) that the intravenous injection to mice of bone marrow not only from mice but from guinea-pigs exerts a protective effect. It is difficult to see how heterospecific cells could act as a seeding material: they could be a source of hormone. Experiments at Harwell (Barnes, 1952) support the contention of Jacobson *et al.* (1951) that implanted mouse spleens exert a measure of protection to mice but so far all heterospecific tissues—guinea-pig or rabbit spleens—have given negative results. Until a cell-free extract has been shown to be effective one would regard the present position as equivocal.

Other procedures which are potentially effective as therapy rather than prophylaxis concern blood transfusion Allen *et al.* (1951) regard the transfusion of whole blood as the most effective agent for the correction of the anaemia and as a source of utilizable protein. With transfusion and aureomycin they obtained a survival of two out of eleven dogs given supralethal doses. Transfusion of fresh whole blood (and possibly other factors) by means of parabiosis had been shown some years ago (Barnes and Furth, 1943) to be protective and lately it has been shown that parabiosis after irradiation is up to a point curative (Brecher and Cronkite, 1951). Early exchange transfusion has also been reported as effective (Salisbury *et al*, 1951).

and Hervé, 1951), methylamine (Hervé, 1951) and β-mercaptoethylamine (Bacq *et al.*, 1951).

The protective procedures may in some cases be due to the induction of a state of anoxaemia—a low oxygen tension is known to be protective to rats and mice (Shack and MacDuffie, 1949; Dowdy *et al.*, 1950) and to plant material (Mottram, 1935; Crabtree and Cramer, 1933). Para-aminopropriophenone induces a methaemoglobinaemia which must lead to under-oxygenation of the tissues. Sodium nitrite may act in a similar way or through acceleration of the decomposition of a catalase—H_2O_2 complex (Chance and Herbert, 1950). Ethanol is a similar accelerator. Sodium cyanide and azide are reducing substances but in the concentrations Bacq was able to use they could not be effective in this way; he suggests that they act through temporary reversible inhibition of certain enzymes. The thiols might act as competitors for oxidizing radicles with the sulph-hydryl enzymes and thus protect them from oxidizing effects of radiation but there is little in the way of objective proof of this. The amines are also protectors of certain enzymes against the effects of the mustard derivatives. Logically one would expect therefore that protection of the integrity of the enzyme systems was the clue to protection of the living animal. However, few enzymes are demonstrably inhibited *in vivo* after doses of radiation of this magnitude Fifty per cent reduction in the pseudo-cholinesterase of the gut has been recorded (Burn *et al*, 1952); inhibition can be shown to occur *in vitro* (Dale, 1947; Barron *et al*, 1949).

By the use of the most effective of these agents (β-mercaptoethylamine) a 95 per cent or more mortality in control mice can be converted to one of less than 10 per cent

Protective biological procedures. In addition to the significant protection induced by anoxaemia or by shielding of the adrenals already mentioned, a number of other prophylactic procedures have been found to be effective.

Treadwell *et al.* (1945) showed that Swiss mice given large doses of oestradiol benzoate about ten days before irradiation were less sensitive. Jacobson *et al.* (1948) showed that rabbits with bone marrow made artificially hyperplastic regenerated

(i) a stage of sickness and shock, possibly due to the circulation of intracellular products of disintegrating cells;

(ii) a latent interval with signs in the peripheral blood of aplasia of the haemopoietic tissues; and

(iii) a final stage where sepsis and bacteriaemia on the one hand and haemorrhage on the other appear superficially to be the immediate cause of death: there are probably, however, more subtle biochemical disturbances which are of greater significance.

In the animal given lethal and supralethal doses, attempts have been made to investigate these more obscure factors. Water and electrolyte balances, the pituitary-adrenal mechanism under this form of stress, and the efficacy of methods of protection by chemical and physiological means are some of the lines which have been extensively studied. With really overwhelming doses of radiation the exposed animal dies with signs of hyperkalaemia.

REFERENCES

Note. U.S. Atomic Energy Commission reports MDDC 250, MDDC 1274, UR 94 and ANL 4676 are published by the Office of Technical Services, Atomic Energy Commission, Washington, 25, D C.

ALLEN, J. C , SANDERSON, M , MILHAM, M., KIRSCHON, A , and JACOBSON, L. O. (1948). *J exp. Med.* 87, 71.

— MOULDER, P. V., and EMERSON, D. M. (1951). *J Amer med Ass.* 145, 704.

BACQ, Z. M., and HERVÉ, A. (1951). *Brit J. Radiol.* 24, 617.

— HERVÉ, A., LECOMTE, J , FISCHER, P , BLAVIER, J., DESCHAMPS, G , [illegible]

(1949). *J. gen Physiol.* 32, 537.

BEAN, W. B., SPIES, T. D., and VILTER, R. W (1944). *Amer. J med Sci.* 208, 46.

BECLERE, A. (1918) *Amer. J. Roentgenol* 5, 498

BENNETT, L. R , REKERS, P. E , and HOWLAND, J W. (1951). *Radiology,* 57, 99.

BETZ, H. (1951). *C.R. Soc. Biol., Paris,* 145, 1240.

THE OVERWHELMING DOSE OF RADIATION

This range of dosage has received very little attention in physiological investigation. Henshaw (1944) exposed mice to an X-ray beam until death occurred within an hour or so after doses of radiation of the order of 30,000r.

A similar type of experiment has been carried out at Harwell using a large source of radiocobalt (Lundie and Barnes, unpublished). Rabbits were exposed and received approximately 15,000r/hour. Most of the rabbits succumbed within 1½ to 3 hours. After about one hour, the rabbits showed incoordinate movements and later apparently convulsions. Blood taken immediately post-mortem showed profound leucopenia and more significantly very high values for serum potassium. Guinea-pigs showed a similar clinical course and when taken out of the beam pre-mortem a state reminiscent of decerebrate rigidity. They lay on their sides and exhibited extension of the hind limbs. The histological findings were not greatly rewarding and again a biochemical lesion seemed the most likely cause of death. The experiments were not continued as the source was only available for a short while.

In continuing the investigation of the hyperkalaemia Mackenzie (unpublished observations) showed that rats irradiated with X-rays at about 1,250r/min began to develop significant increases in serum-potassium after 40,000r and similar increases in the red cell sodium. The animals died with electrocardiographic changes compatible with potassium intoxication after 80,000 to 100,000r.

SUMMARY

In the past X- and γ-radiations have been widely used for local destruction of unwanted tissue such as new growths.

Today with increased use of nuclear energy one requires to know how radiation given to the whole body affects its normal physiological and biochemical processes. So far experimentation with laboratory animals has been largely confined to the acute disturbances following large doses.

In the region of the median lethal dose the 'radiation syndrome' consists of three phases.

HUNTER, C. G. (1952). Personal communication.
IMLER, A. E., and WAMMOCK, H. (1940). *Amer. J. Roentgenol.* **43**, 243.
INTERNATIONAL PROTECTION COMMISSION (1951). *Brit. J Radiol.* **24**, 46
JACOBSON, L O., MARKS, E. K, SIMMONS, E. L, HAGEN, C. W, and ZIRKLE, R. E. (1947). MDDC 1174
— MARKS, E. K., GASTON, E. O., SIMMONS, E. L., and BLOCK, M. H (1948). *Science*, **107**, 248.
— MARKS, E. K, ROBSON, M. J., GASTON, E. O., and ZIRKLE, R. E (1949). *J Lab clin Med.* **34**, 1538.
— SIMMONS, E. L, MARKS, E. K, GASTON, E. O., ROBSON, M J., and ELDREDGE, J. H. (1951). *J. Lab clin Med* **37**, 683.
JENKINSON, E. L, and BROWN, W. H. (1944) *Amer. J. Roentgenol.* **51**, 486.
LAWRENCE, J. S., DOWDY, A. H., and VALENTINE, W. N. (1948) *Radiology*, **51**, 400
LEA, D E. (1946) *Actions of Radiations on Living Cells* Cambridge University Press
LEBLOND, C. P., and SEGAL, G. (1942). *Amer. J. Roentgenol.* **47**, 302.
LE ROY, G V. (1950). *Arch. intern. Med* **86**, 691.
LIEBOW, A. A., WARREN, S., and DE COURSEY, E. (1949). *Amer J Path* **25**, 853.
LIMPEROS, G, and MOSHER, W. A. (1950). *Science*, **112**, 86.
LOFSTROM, J. E., and NURNBERGER, C E (1946) *Amer. J. Roentgenol.* **56**, 211
LORENZ, E, UPHOFF, D, and CONGDON, C. (1951) *Argonne Nat Lab. Div. of Biol and Med. Res* ANL 4676, p. 14
— UPHOFF, D, REID, T. R, and SHELTON, E. (1951) *J. Nat Cancer Inst* **12**, 197

— HAMMOND, C W., and TOMPKINS, M (1950b) *Science*, **111**, 719.
MOLE, R H., PHILPOT, J. ST L., and HODGES, G R. V. (1950) *Nature, Lond* **166**, 515
MOTTRAM, J C. (1933) *Nature, Lond* **132**, 317
— (1935) *Brit. J Radiol.* **8**, 32.
PAINTER, E. E, PROSSER, C L, and MOORE, M C (1946) MDDC 661.
PATERSON, E., and MATTHEWS, J J. (1951) *Nature, Lond* **168**, 1126
PATT, H M., SWIFT, M. N, TYREE, E B, and JOHN, E S (1947) *Amer J. Physiol* **150**, 480.
— TYREE, E. B, STRAUBE, R. L, and SMITH, D E (1949) *Science*, **110**, 213
— SMITH, D E, TYREE, E. B, and STRAUBE, R L (1950). *Proc Soc exp Biol, N.Y.* **73**, 18.
PENICK, G. D, CRONKITE, E P, GODWIN, I D, and BRINKHOUS, K M (1951) *Proc Soc exp. Biol., N.Y* **78**, 732.

BLOOM, W. (1948). *Histopathology of Irradiation.* McGraw Hill, New York.
BOCHE, R. D., and BISHOP, F. W. (1946). MDDC 250.
BOWERS, J. Z , and SCOTT, K. G. (1951). *Proc. Soc. exp. Biol. and Med* **78**, 645 and 648.
BRECHER, G., and CRONKITE, E. P. (1951). *Proc. Soc. exp Biol and Med.* **77**, 292.
BURN, J. H., KORDIK, P., and MOLE, R. H. (1952). *Brit J. Pharmacol* **7**, 58.
CASPARI, W. (1928). 'Biologische Grundlagen der Strahlenbehandlung besartiger Geschwulste' in *Handbuch d. ges. Strahlenheilkunde*, vol 1, p. 771 Ed. P Lazarus Bergmann. Munich.
CHANCE, B , and HERBERT, D. (1950). *Biochem. J.* **46**, 402.
CHAPMAN, W. H., SIPE, C R., ELTZHOLTZ, D. C., CRONKITE, E. P , and CHAMBERS, F. W. (1949). Naval Med. Res. Inst. Project NM. 006 012.08.25
COLE, L. J., BOND, V. P., and FISCHLER, M C. (1951). U.S. Nav. Radiol. Defense Lab. Report AD-331 (B).
COURT-BROWN, W M. (1948). *Brit J Radiol.* **21**, 221.
CRABTREE, H. G , and CRAMER, W. (1933) *Proc. Roy. Soc. B.* **113**, 226.
CRONKITE, E P (1949). *Mil. Surg* **104**, 7.
— ELTZHOLTZ, D C , SIPE, C. R., CHAPMAN, W. H., and CHAMBERS, F. W. [illegible]
— [illegible]
— [illegible] W. (1950) *Proc. Soc. exp. Biol , N Y.* **73**, 184.
— JACOBS, C. J , BRECHER, G., and DILLARD, G H. L. (1952) *Amer J. Roentgenol.* **67**, 796.
DALE, W. M. (1947) *Brit. J Radiol.* Supp. 1 46.
DOWDY, A. H , BENNETT, L R , and CHASTAIN, S. M. (1950). *Radiology*, **55**, 879
DUGGAR, B. M (1936) *Biological Effects of Radiation.* McGraw Hill, New York.
EDELMANN, A (1951). *Amer J Physiol* **165**, 57.
— and EVERSOLE, W J (1951) *Amer J. Physiol.* **163**, 709
ELLINGER, F. (1945) *Radiology*, **44**, 241
— (1948) *Radiology*, **51**, 394
— ROSWIT, B , and GLASSER, S M (1949). *Amer. J Roentgenol* **61**, 387.
ELLIS, F. (1942) *Brit J Radiol* **15**, 194.
FIELD, J B., and REKERS, P E (1949) *J clin. Invest.* **28**, 746
FURTH, F W , COULTER, H P., and HOWLAND, J. W. (1952). *Amer. J. Path.* **28**, 25.
[illegible]
[illegible] ER, M , and McDONNEL, G M (1949). University of Rochester, UR 94.

Plate I

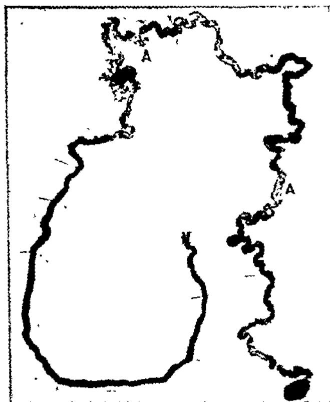


Fig. 19 Effects of haemorrhagic shock on human proximal tubule Copious haemorrhage in third stage of labor. 'The greater part of the proximal tubule is well preserved, the heavy staining was deliberately done to bring out the contrast of the two regions marked A, where stretches of the convolution show damage.' In the more distal lesion there is complete disruption of the tubule wall with surrounding tissue reaction From Oliver, MacDowell, and Tracy (1951)

(*See p 185*, 'Studies of Physiology of the Kidney')

Prosser, C. L. (1947). *Radiology*, 49, 299

Rigdon, R. H., and Curl, H. (1943). *Amer. J. Roentgenol.* 49, 250.

Salisbury, P. P., Rekers, P. E., Miller, J. H., and Marti, N. F. (1951). *Science*, 113, 6.

[illegible]

Soc. exp. Biol., N.Y. 73, 198.

Smith, W. W., Smith Falconer, and Thompson, E. C. (1950). *Proc. Soc. exp Biol., N.Y.* 73, 529.

Spear, F. G (1946). *Brit. med. Bull.* 4, 2.

[illegible] 202 [illegible] *iol, N.Y.* 74, 242.

— Moshman, J., and Furth, J. (1951). *Science*, 114, 665.

Treadwell, A. de G., Gardner, W. N., and Lawrence, J. H. (1943). *Endocrinology*, 32, 161.

Trowell, O. A. (1952). *J. Path. Bact* 64, 687.

U.S. Atomic Energy Commission and Department of Defense (1950). *The Effects of Atomic Weapons*, McGraw Hill, New York.

Warren, S. L., and Whipple, G. H. (1923). *J Amer. med. Ass.* 81, 1673.

Warren, S., and Bowers, J. Z (1950). *Ann. intern Med.* 32, 207.

Weisberger, A. S, Heinle, R. W., Storaasli, J. P., and Hannah, R. (1950). *J. clin. Invest* 29, 336.

Plate III

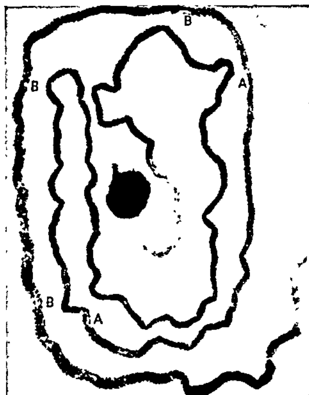


Fig 21 Proximal tubule and glomerulus from dog's kidney taken 5 days after the renal artery had been clamped for 66 minutes 72 minutes after removal of the clamp the dog received 50 cc. of the dog's red blood cells laked in distilled water, with salt added to isotonicity and 50 cc. of the animal's plasma

The part of the tubule next the glomerulus shows complete disintegration of the basement membrane and disruption of the tubule wall The convolution from here to A is well preserved. At A the mitochondrial pattern is disturbed and at B the basement membrane is disintegrated and the cellular detail lost There was no evidence of toxic damage, despite the infusion of haemoglobin From Oliver, MacDowell, and Tracy (1951)

(*See p. 185*, 'Studies of Physiology of the Kidney')

PLATE II

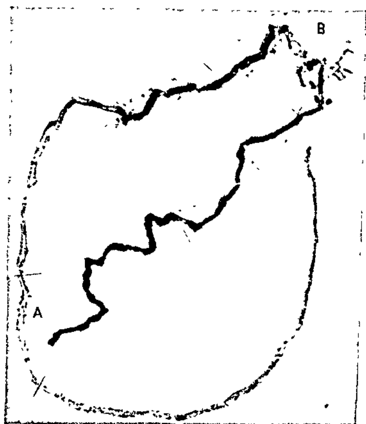


FIG 20 Ascending limb and distal convolution from the same kidney

the case, there is no apparent lesion in this tubule in the proximity of the pigment casts From Oliver, MacDowell, and Tracy (1951)

(See p 183, 'Studies of Physiology of the Kidney')

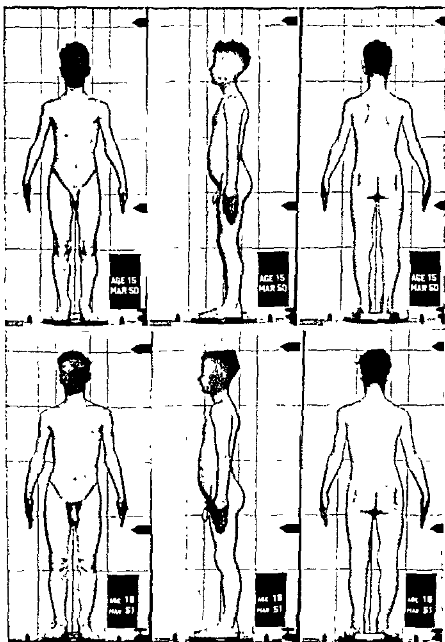


FIG. 15. Photogrammetric pictures of a child.

(*See p 342*, 'Growth of the Human at the Time of Adolescence')

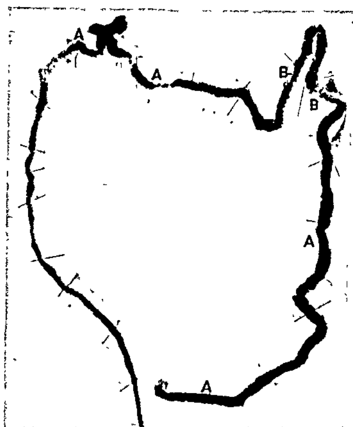


Fig. [illegible] Ascending limb and distal convolution from the same kidney. [illegible] ed A the [illegible] positive [illegible] At places marked B there is disintegration of the basement membrane and a complete disruption of the tubule wall At upper B crystals of haeme pigment are present From Oliver, MacDowell, and Tracy (1951)

(*See p 185*, 'Studies of Physiology of the Kidney')

PLATE VI



FIG 17. Technique used for soft tissue radiographs of calf, thigh, and arm Left, to show positioning in regard to tube and casette; right, to show measurement of distance from central plane of limb to film.

(*See p. 343*, 'Growth of the Human at the Time of Adolescence')